中国数学发展的若干主攻方向

Some Main Aspects on Development of Mathematics in China

程民德 主编

Editor in Chief Cheng Minde

江苏教育出版社

中国数学发展的若干主攻方向

Some Main Aspects on Development of Mathematics in China

Editor in Chief
Cheng Minde





程民德 主编

（苏）新登字第003号

中国数学发展的若干主攻方向

程民德 主编
责任编辑 喻伟

出版发行：江苏教育出版社
（南京中央路165号，邮政编码：210009）
经 销：江苏省新华书店
印 刷：阜宁人民印刷厂
（阜宁大街173号 邮政编码：224400）

开本850×1168毫米 1/32 印张14.75 插页4 字数359，000
1994年10月第1版 1994年10月第1次印刷
印数1 2，000

ISBN 7—5343—2172—7

G·1934 定价：11.10元

程民德
（北京大学数学研究所）
Cheng Minde
(Institute of Mathematics,
Peking University)

序
Preface

本文集所列十项主攻方向，可以追溯到1988年1月由国家自然科学基金委员会组织实施的“七五”数学重大项目．该重大项目的名称是《现代数学中若干基本问题的研究》，它的十个课题就是本文集的十项主攻方向，这并不是偶然的巧合．而是事物发展后我们在认识上的进展．本文集早在1988年8月由当代数学大师陈省身教授发起并主持的“21世纪中国数学展望学术讨论会”在南开数学研究所举行第一次会议时，江苏教育出版社的同志就和我们酝酿征稿出书的事，目的在介绍该重大项目的十个课题，使我国科学界和教育界有更多的人了解现代数学发展中的一些重要课题，引起社会上对数学的重视，促使我国数学率先赶上世界先进水平．当时已按课题名称开始征集介绍文章，文集的名称也暂定为《现代数学中的若干基本问题》，在提法上和该重大项目的名称相同．当时我们的认识虽已明确该重大项目的十个课题是现代数学中的一些基本而重要的问题，至于它们能否都成为现代数学中的主攻方向，我们认为还有待于项目进展以后的具体情况来考虑．国际上不乏这样的先例，有些方向热闹一时，大家工作一段时间后，就发现已没有多大发展余地了．1991年5月在南开数学研究所举行了第二次“21世纪中国数学展望学术讨论会”，正值“七五”数学重大项目由国家自然科学基金委员会组织总结验收，该项目的十个课题都超额完成了预期目标，取得了丰硕成果．更可喜的是每个课题都提出了具有重要意义的新问题，并都被列入相应的“八五”数学重点项目之中．这一事实充分说明“七五”数学重大项目所列的十个课题是富有生命力的，它们在超额完成了“七五”所提的指标后，又将在“八五”期间向新的更高的指标迈进，称它们为现代数学中的若干主攻方向，现在看来是完全允当的．根据这样的认识，本文集才正式定名为《中国数学发展的若干主攻方向》，同时本文集不仅是对“七五”数学重大项目的十项主攻方向作介绍，也是对“八五”数学部分重点项目的主攻方向作介绍．这一进展谅为读者所欢迎．

"七五"数学重大项目的十个课题能成为现代数学中的若干主攻方向也并不是偶然的,这可以从现代数学的发展趋势和当时课题设置的主导思想来分析.在酝酿"七五"数学重大项目的课题设置时,我们曾多次征询了工作在数学研究第一线的许多数学界的同仁们的意见,经过反复认真的讨论,一致认为课题设置不宜仅限于学科分支来考虑,应着眼于现代数学发展的趋势.一些在现代数学发展中富有生命力的重要学科分支如"代数几何与代数数论"、"动力系统"、"复分析"等自应列为课题,它们有较强的综合性,已不是单个的学科分支.但更应考虑到现代数学发展趋势的特点,即数学各分支学科的相互渗透以及数学与其他学科的相互联系.一些重要发展方向往往是跨学科分支甚至跨越整个数学学科的,很难归入某一个传统的学科分支,例如课题中的"计算机数学"和"非线性分析"就是如此,它们是重要发展方向,不仅是跨学科分支的,它们的应用范围更是跨越数学学科的.另外如"整体微分几何及其物理应用(兼含低维拓扑)"、"粒子系统与随机分析"和"偏微分方程数值解法及在科学、工程中的应用"这几个课题的名称就显示了它们是跨越数学学科的.我们课题设置的另一主导思想是理论与应用相互促进,理论课题一般有具体的实际背景和宽广的应用前景.应用课题如"最优化、辨识与控制"和"多维数据分析与统计推断",都有精深的理论根据和推理分析,在"七五"数学重大项目的总结中,我们已历举了各课题在理论和应用方面突出的成就,显示了理论和应用的相互促进.正由于我们对"七五"数学重大项目的课题设置能着眼于现代数学的发展趋势,并能注意到理论和应用的相互促进,从而这十项课题能成为现代数学的主攻方向只不过是它们在正常进展中的自然后果.

关于现代数学的发展趋势,我们还可以从1990年8月在日本京都举行的第21届国际数学家大会颁发的Fields 奖得到一些启发.这次大会颁发了四个Fields 奖,它们的得主是美国的E. Witten,

新西兰的 V. F. R. Jones、俄罗斯的 V. G. Drinfeld 和日本的森重文.前三人的获奖工作都与近代理论物理学的发展有不同程度的联系,特别是 Witten 的工作物理味道更浓,只有森重文的工作是属于代数几何这一数学分支学科,他的三维代数几何新成就联系着许多过去在这分支学科方面获得 Fields 奖者的工作,他的获奖是在人意料中的.这次颁奖情况反映了数学研究与物理学的联系有明显的优势,已引起国际数学界的注目.数学与物理学是两门有悠久发展历史的学科,它们之间的密切联系使这两门学科都各自得到了巨大的发展,现在看来这种深刻的内在联系,不论这两门学科各自发展得多远,依旧在起着极为奥妙而持久的作用.另外从这前三人的获奖工作来看,也生动反映了不同学科之间的相互交叉. Jones 从纽结理论获得了他的著名多项式,Witten 从量子力学中粒子的状态空间的理论重新发现了 Jones 多项式;“辫群”理论进入了 Drinfeld、Jones 和 Witten 三人的不同工作之中;追溯到上届 Fields 奖获得者 Donaldson 的工作,它反映出来源于物理学的杨-Mills 方程进入了4维流形的研究.凡此等等都反映出现代数学发展趋势的一些特点.

关于“七五”数学重大项目所取得的科研成果是十分丰富多采的,其中获奖的科研工作就有36项,比较重要的奖有国际性的第三世界数学奖1项,国家自然科学奖三等奖4项,全国数学陈省身奖3项,科学院自然科学一等奖6项.项目的科研成果在国内外重要学术刊物上发表的论文有693篇,在国际重要学术会议上所作的特邀报告和大会报告有50个.所有十个课题无一例外地都取得不少具有国际先进水平的科研成果.其中“计算机数学”课题在数学机械化研究中形成的关于非线性代数方程组的高效算法,以其理论的完整和应用的普遍,引起国际学术界和高技术科学界的广泛重视,并通称之为吴方法;“偏微分方程数值解法及在科学、工程中的应用”课题中关于 Hamilton 系统的辛几何算法以其创新性、系统性

和成果的丰富多采受到国际同行著名数学家的重视;这两项工作是我国独创的并已引发了国内外一系列后继研究和应用,是属于国际领先水平的研究领域.本文集只是对这十个课题的一些研究方向作概括性的介绍,当然不可能一一涉及各课题所取得的重要科研成果.另一方面本文集虽力图能比较全面地介绍出这十项主攻方向来,江苏教育出版社在征集稿件方面确也作了很大的努力,但各课题征得的稿件依旧是很不平衡的.例如在十个课题中,以"偏微分方程数值解法及在科学、工程中的应用"这一课题的规模最大,它具有较多的子课题,从而在"七五"期间取得的成果也最多.不仅如此,它在"八五"重大项目"大规模科学与工程计算的方法和理论"中有了继承,同时它还有一些子课题准备列入"八五"的重点项目.但本文集对这规模最大的主攻方向只有一篇介绍,自然很难反映这一主攻方向的全貌.至于"计算机数学"课题的情况更是如此,它本以数学定理机器证明及其应用为主要内容,在"八五"重大项目"机器证明及其应用"中有了继承,而且该重大项目已列入国家科委的"攀登计划",但对这一最主要的主攻方向,本文集也只有一篇介绍,同样很难反映出它的全貌.这个缺陷对其他主攻方向也同样存在,只是程度不同罢了,希望读者能予谅解.上文中提到的"八五"数学重点项目,指的是国家基金委组织的资助强度仅次于重大项目的科研项目,而"八五"重大项目一般是跨学科的,不再设单学科的重大项目.目前已论证过的"八五"重大项目中与数学直接有关的共有三项,其中两项是上述的"大规模科学与工程计算的方法和理论"和"机器证明及其应用",另一项是"非线性科学".在这篇短短的序言中也没有可能把"七五"数学重大项目的主要收获一一描绘出来,我愿把该项目的验收评议书的最后一句话摘录于下:"将本项目作为重大项目给予较强的资助,促进我国数学事业的发展进程,为实现使我国数学率先赶上世界先进水平作出了贡献,是一个成功的经验".我认为这是对"七五"数学重大项

目整个组织实施后取得的成绩作了一个总的概括.

本文集的出版得到国家自然科学基金委员会数理科学部的支持和“七五”数学重大项目各课题负责人的合作,在此一并表示由衷的感谢.江苏教育出版社自始至终为出好本书进行了不懈的努力,也是为实现使我国数学率先赶上世界先进水平作出了贡献,我们表示深切的感谢.文集的编写上难免有不当之处,希读者不吝赐教!

程民德 于北京大学数学研究所

1993年11月28日

目　　录

《中国数学发展的若干主攻方向》序

I　计算机数学

II　非线性分析

III　动力系统

IX 粒子系统与随机分析

X 多维数据分析与统计推断

八五数学重点项目简介

Contents

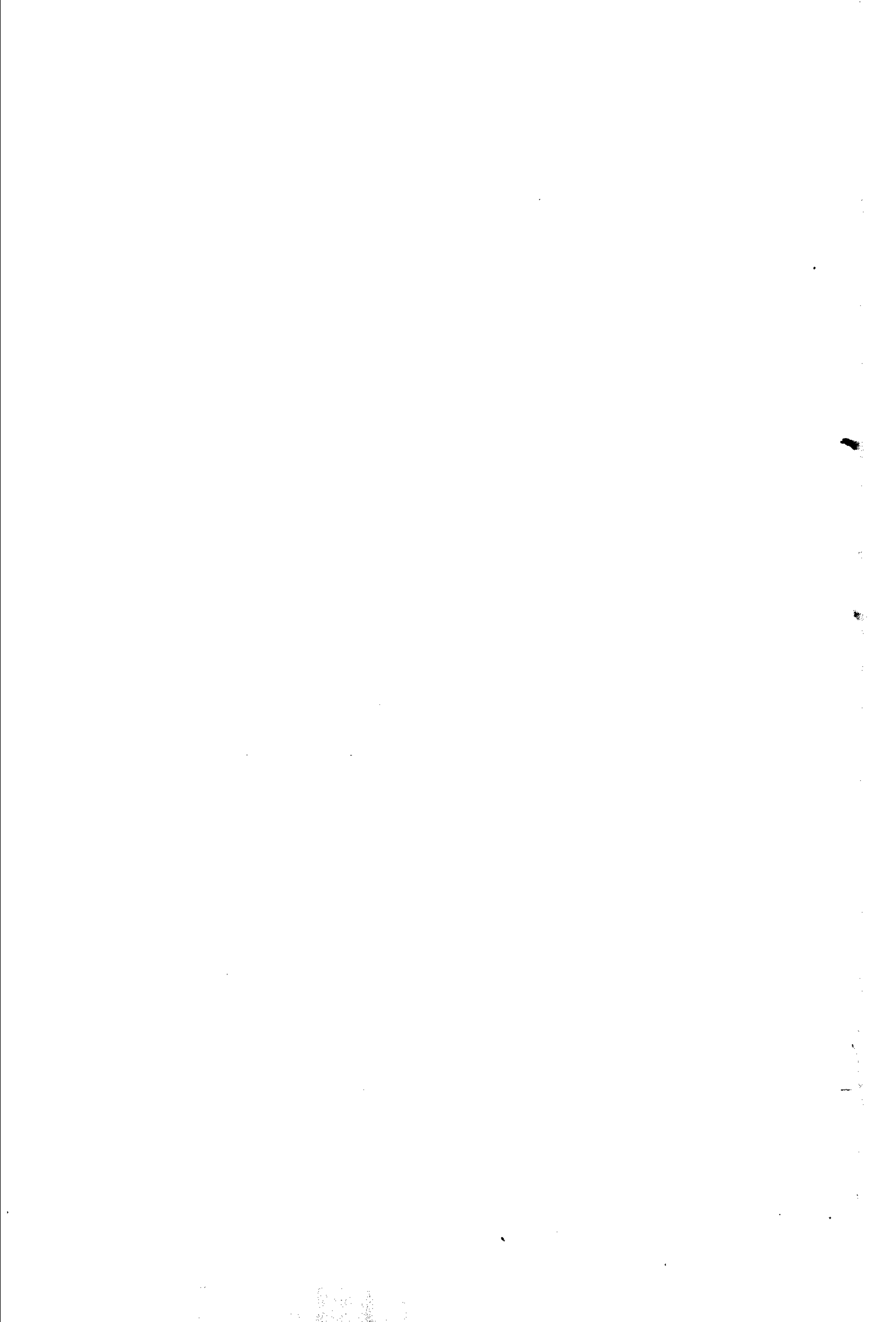
VII Algebraic Geometry and Algebraic Number Theory

VIII Optimization, Identification and Control

IX Interacting Particle Systems and Stochastic Analysis

I

计算机数学
Computer Mathematics



石　赫
（中国科学院系统科学研究所）
Shi He
(Institute of Systems Science,
Academia Sinica)

数学定理机械化证明的吴文俊原理
Wu Wentsun Principle of Mechanical Proving on Mathematical Theorems

所谓数学定理之证明，无非是从定理的假设出发，以数学公理为依据，利用已知的数学定理，经过一步一步演绎推理，来确证定理的结论为真. 建立在公理化体系上的数学典范首推欧几里德(Euclid)几何. 约公元前3世纪问世的《几何原本》，以“每理必证”为其特色. 大家知道，欧氏几何中的定理无穷无尽，而这些定理的证明方法却又“因理而异”—— 定理不同，其证明方法亦不同. 它们各具巧思，各有特色，难以发现其规律. 中学里设置的几何课程，这种“一理一证”的特点给初学者带来何其多的烦恼，而他们在绞尽脑汁冥思苦索后有所发现之时，又会感受到“苦尽甘来”的欣喜. 正是在这种磨炼之中，学生们领略了数学发展的历史过程，接受了数学思维和基本技能的训练，为以后的学习和研究奠定了必要的基础.

何谓数学定理的机械化证明？一言以蔽之：“万理一证”. 即把某一类定理(这类定理可能成千上万，也可能无穷无尽)当做一个整体加以考虑，建立一种统一的、确定的证明程序，使得该类中的每个定理，只要按照此程序机械地(死板地)、按部就班地一步一步实施下去，经有限步之后即可从数学命题的假设推断出数学命题的结论是否为真. 命题为真时即为定理. 对一类定理而言，这种统一的、确定的证明方法称为定理证明的机械化方法. 依照机械化方法实施数学定理的证明称为数学定理的机械化证明. 建立和提供数学定理证明机械化方法的数学原理称为数学定理证明的机械化原理. 自然，从“一理一证”到“万理一证”，即实现数学定理证明的机械化，是数学的认识和实践中的一次飞跃.

数学机械化的思想曾吸引了无数数学家为之倾注心血. 建立坐标系从而把变量引入数学的笛卡尔(Descartes)在其未发表的著作中提出过宏伟的设想：一切数学问题都化为代数问题，而一切代数问题都归结为单个代数方程的求解问题. 尽管笛卡尔的设想未能实现，但这一深刻的思想对其后的数学发展有着重大影响. 微

积分学的先驱莱伯尼兹(Leibniz)设想过数学领域的推理机器，这可看作机械化思想的初步尝试．对20世纪数学发展作出重大贡献的希尔伯特(Hilbert)，在其名著《几何基础》中强调了数学公理化体系的建立．也就是在这本著作中，Hilbert同时叙述了数学机械化的思想，写进了数学著作中最早出现的一条机械化定理．[2]中称其为Hilbert机械化定理．Hilbert本人在其数十年的数学生涯中，不断探求数学机械化的途径．然而，Hilbert对数学的这一深刻认识却长期为数学家们所忽视．冯·诺伊曼(von Neumann)自然是数学机械化的积极倡导者．数学机械化的研究以及电子计算机的出现，使数理逻辑这门学科应运而生．可是当它获得长足的发展之后，却把其热情倾注于计算机科学，小心翼翼地同数学的机械化保持一段距离．这些数学大师们以及无数数学家坚持不懈地求索，表明了数学机械化的思想重要而深刻；而他们遭受的挫折，同时也意味着数学机械化的道路漫长而艰难．

数学定理的机械化证明须要构造性数学．对机械化数学而言，只证明数学问题的解的存在性是远远不够的，必须有具体的、确定的算法把解构造出来．当数学问题的解具体构造出来之时，它的存在性也就不证自明了．就此而言，构造性数学要比存在性数学困难得多．在数学发展的历史长河中，如果说古希腊的数学及其延续以数学公理化及演绎推理为其特点，那么从问题出发，建立算法的机械化则是古代中国数学研究的传统．成书于公元前的《九章算术》是算法机械化的光辉典范．《九章算术》归纳了九类数学问题，每章中讨论一类问题，给出解答，并具体讲明该类问题的求解方法，算法成立的理由则在注释中加以说明(见[5])．从问题出发，“寓理于算，不证自明”是该书的特色．《九章算术》中明确给出了解线性方程组的消元法，这要比高斯(Gauss)早1800多年．被西方推崇为中世纪最伟大的数学家的朱世杰(见[8])，他撰著的数学名著《四元玉鉴》于1303年出版．在这部数学经典中已给出了高次代数联

立方程组的解法，未知数的个数最多可达 4 个. 而这类问题的求解却长时期困惑着近代数学家们. 合理地继承和发扬中国古代数学研究的创造思想，不仅是数学史学家们的任务，而且也应当是每位中国数学工作者的责任. 吴文俊机械化原理的创立，从思维到方法都受益于中国古代数学研究的构造性思想.

前面讲到的 Hilbert 机械化定理断言：一类纯交点型定理可以实现机械化证明. 这类定理要适合如下的假定：定理中设定的变元应以线性方式出现在代数关系式中，所以这类定理只能描述点和直线的从属关系，诸如：点在线上，线过一点，二线交于一点，过已知点作一直线平行于给定直线，等等. 初等几何中最基本的概念“距离”以及最基本的图形“圆”尚未能包括在内. 这就使得依 Hilbert 机械化定理能实现机械化证明的数学定理的范围过于狭窄. 1950 年，塔斯基(Tarski) 证明，实数闭域上的初等几何问题可以机械化证明. 但是他所给出的方法极为繁难，虽经多人改进，时至今日仍然不能具体实现. 只是在中国数学家吴文俊先生创立了数学定理机械化证明的吴氏原理之后，人们才第一次得以实现大量相当困难的数学定理的机械化证明.

1977 年，吴文俊先生关于数学定理机械化证明的重要论文《初等几何判定问题与机械化问题》发表. 1984 年，学术专著《几何定理机器证明的基本原理》(以下简称为《机证原理》) 由科学出版社出版. 这本专著遵循机械化思想引进数系和公理，依照机械化观点系统地分析了各类几何体系，诸如 Pascal 几何，垂直几何，度量几何，以及欧氏几何(即常用几何)，明确建立了各类几何的机械化定理.《机证原理》一书，系统地阐明了几何定理机械化证明的基本原理，奠定了数学机械化体系的基础.

几何问题的代数化是几何问题机械化的第一步. 为此需要引进数系，建立坐标系，把几何构图中的各种关系利用代数方程来描述. 在适当选取坐标系之后，如果数学定理的假设条件可表示为一

组代数方程

$$(H):\begin{cases} f_1(x_1,x_2,\cdots,x_n)=0;\\ f_2(x_1,x_2,\cdots,x_n)=0;\\ \cdots\\ f_r(x_1,x_2,\cdots,x_n)=0. \end{cases}$$

而数学定理的结论由代数方程

$$C:g(x_1,x_2,\cdots,x_n)=0$$

所刻画，这里$f_1,f_2,\cdots,f_r$和g都是变元$x_1,x_2,\cdots,x_n$的多项式，那么数学定理的机械化证明就归结为如下的问题.

机械化问题 构造并提供一种确定的、机械的算法，使得依此算法进行有限步之后即可判定：在若干附加条件之下，结论C是否可由假设H推出，即是否可由$f_1=0,f_2=0,\cdots,f_r=0$推出$g=0$.

数学定理机械化证明的吴氏原理完满地解答了这一机械化问题.

多项式组的升列和不可约升列的概念是吴氏原理中的基本概念. 设K为特征为零的数域(比如，有理数域)，K上以变元$x_1,x_2,\cdots,x_n,u_1,\cdots,u_r$为不定元的多项式环记为$K[x_1,x_2,\cdots,x_n,u_1,\cdots,u_r]$，其中$x_1,x_2,\cdots,x_n$称为约束变元，$u_1,u_2,\cdots,u_r$称为自由变元.

定义 如果多项式组$A_1,A_2,\cdots,A_n\subset K[x_1,\cdots,x_n,u_1,\cdots,u_r]$，满足下列条件，则称其为升列.

i) 对任一多项式A_i，A_i中包含的变元的最大下标为i，即A_i可按x_i的幂次写为

$$A_i(x_1,\cdots,x_i,u_1,\cdots,u_r)=I_ix_i^{m_i}+x_i\text{ 的低次项};$$

其中I_i为$K[x_1,x_2,\cdots,x_{i-1},u_1,\cdots,u_r]$中的多项式，称为$A_i$的初式.

ii) 每一A_i，$i=2,3,\cdots,n$，对任意的$j<i$，A_i对A_j已经约化. 即A_i的初式I_i中出现的x_j幂次要低于A_j中x_j的幂次m_j.

还需要把升列的定义稍许扩充：

定义 如果升列中包括有只含自由变元 $u_1,\cdots,u_r$ 的多项式 A_0，则称该升列是矛盾的.

定义 如果升列 $A_1,A_2,\cdots,A_n$ 满足下列条件，则称其为不可约升列.

i） 记 $K_0=K(u_1,\cdots,u_r)$ 为数域 K 添加未定元 $u_1,u_2,\cdots,u_r$ 的扩域. A_1 作为 K_0 上的多项式是不可约的；

ii） 一般而言，对每个 $i,i=2,3,\cdots,n$，设 η_{i-1} 为多项式 A_{i-1} 的零点，记 $K_{i-1}=K_{i-2}(\eta_{i-1})$ 为 K_{i-2} 添加 η_{i-1} 的扩域. A_i 作为 K_{i-1} 上的多项式是不可约的.

从上述定义可以看出，升列具有如下的规范的"三角化"形式：

$$
\begin{aligned}
A_1 &= A_1(x_1,u),\\
A_2 &= A_2(x_1,x_2,u),\\
&\cdots\\
A_{n-1} &= A_{n-1}(x_1,x_2,\cdots,x_{n-1},u),\\
A_n &= A_n(x_1,x_2,\cdots,x_n,u),
\end{aligned}
$$

其中 $u=(u_1,u_2,\cdots,u_r)$.

设给定升列 $A_1,A_2,\cdots,A_n$，其中 A_i 的初式记为 $I_i,i=1,2,\cdots,n$. 则对于多项式环 $K[x_1,x_2,\cdots,x_n,u_1,\cdots,u_r]$ 中的任一多项式 $G(x_1,x_2,\cdots,x_n,u_1,\cdots,u_r)$，都可利用此升列尽可能降低 G 中约束变元的幂次. 首先 G 对 A_n 求余式 R_{n-1}，R_{n-1} 中 x_n 的幂次低于 A_n 中 x_n 的幂次 m_n，然后 R_{n-1} 对 A_{n-1} 求余式 R_{n-2}，R_{n-2} 中 x_{n-2} 的幂次低于 A_{n-1} 中 x_{n-2} 的幂次. 依次进行下去，最后可得

$$I_1^{s_1}\cdots I_n^{s_n}G=Q_1A_1+Q_2A_2+\cdots+Q_nA_n+R_0,$$

其中 $s_1,s_2,\cdots,s_n$ 为非负整数，$Q_1,Q_2,\cdots,Q_n$ 为 $x_1,\cdots,x_n,u_1,\cdots,u_r$ 的多项式. 这一公式称为 G 对升列 $A_1,A_2,\cdots,A_n$ 的余式公式，R_0 称为 G 对 $A_1,A_2,\cdots,A_n$ 的余式. 须要指出：余式 R_0 一般仍为变元 $x_1,x_2,\cdots x_n,u_1,\cdots,u_r$ 的多项式，只是 R_0 中约束变元 $x_i(i=1,2,$

$\cdots,n$)的幂次低于 A_i 中 x_i 的最高幂次 m_i.

根据余式公式,可以获得数学定理机械化证明的基本定理.

基本定理(吴文俊) 给定升列 $A_1,A_2,\cdots,A_n$,其中 A_i 的初式为 $I_i,i=1,2,\cdots,n$. 对任意的多项式 G,如果 G 对升列 $A_1,A_2,\cdots,A_n$ 的余式 R_0 为零,则在附加条件 $I_i\neq 0,i=1,2,\cdots,n$ 之下,可由 $A_1=0,A_2=0,\cdots,A_n=0$ 推出 $G=0$. 如果升列 $A_1,A_2,\cdots,A_n$ 是不可约的,则余式 R_0 为零也是 $G=0$ 可从 $A_1=0,A_2=0,\cdots,A_n=0$ 推出的必要条件.

这一基本定理为实现数学定理的机械化证明开辟了有效的途径,然而,基本定理是对升列而言的.升列是一种具有规范形式的多项式组,其中排列有序的约束变元是依次出现的.一般情形下,根据数学定理的几何构图所获取的代表数学定理假设条件的多项式组,其中变元出现的状况往往杂乱无章前后交错.因此,如何把杂乱无章的多项式组整顿得井然有序,即从一般的多项式组获取相应的升列,是实现数学定理机械化证明的关键.

现假设给定多项式组 PS:$p_1,p_2,\cdots,p_d$,其中每一多项式是变元 $y_1,y_2,\cdots,y_m$ 的多项式.对变元 $y_1,y_2,\cdots,y_m$ 适当排定次序,则有

整序定理(吴文俊) 存在机械化算法,使对任意给定的多项式组 PS:$p_1,p_2,\cdots p_d$,依此算法机械地进行有限步之后,或得一矛盾列,或得一非矛盾升列

$$\mathrm{CS}:A_i=A_i(x_1,x_2,\cdots,x_i,u_1,\cdots,u_r)=I_ix_i^{m_i}+x_i \text{ 的低次项},$$
$$i=1,2,,\cdots,n,$$

其中 $x_1,x_2,\cdots,x_n,u_1,\cdots,u_r$, $n+r=m$,为变元 $y_1,y_2,\cdots,y_m$ 的某一排列.

升列CS称为多项式组PS的特征列.从多项式组PS推导出特征列CS的过程称为对多项式组PS进行整序.整序过程是按确定的机械的步骤进行的.整序过程中变元 $y_1,y_2,\cdots,y_m$ 自动区分为

约束变元和自由变元. 多项式组 PS 和相应的特征列之间的代数关系是:

i) 多项式组 PS 中任一多项式 p_i 对特征列的余式为零.

ii) 如果多项式组 PS 的公共零点集(即由 $p_1 = 0, p_2 = 0, \cdots, p_n = 0$ 所定义的代数簇) 记为 Zero(PS), 则特征列 CS 中每一多项式 A_i 在 Zero(PS) 上为零.

设 $I = I_1 \cdot I_2 \cdot \cdots \cdot I_n$ 为特征列中诸初式之乘积. 命 Zero(CS/I) 表示 CS 的公共零点集中那些不使 I 为零者, 并且命 $PS_i = \{p_1, p_2 \cdots, p_n, I_i\}$ 为把初式 I_i 添加到多项式组 PS 中得到的扩大的多项式组, 那么, 多项式组 PS 和相应的特征列组 CS 的零点集之间的关系由下列定理描述.

零点结构定理(吴文俊) 多项式组 PS 和经整序后得到的特征列 CS 为已知, 则有

$$\mathrm{Zero(PS)} = \mathrm{Zero(CS/I)} + \bigcup_i \mathrm{Zero}(\mathrm{PS}_i),$$

其中诸符号如前面所定义.

这一定理精确地刻画了多项式组和它的特征列之间的关系, 保证了由特征列替代多项式组, 即将数学定理假设条件所获得的多项式组替换为由它经整序得到的特征列的合理性.

须强调的是: 特征列 CS 中诸初式 $I_1, I_2, \cdots, I_n$ 是在对多项式组 PS 整序过程中自动产生的, 它们是由多项式组 PS 本身所确定的. 如果多项式组 PS 代表数学定理的假设条件, 则 $I_i = 0$ 往往表示了数学定理本身的某种退化, 或者说假设本身的某种不合理性. 因此, $I_i \neq 0, i = 1, 2, \cdots, n$, 可称为数学定理成立的非退化条件. 数学定理机械化证明的基本定理表明, 从数学定理的假设推出数学定理的结论, 须要假定非退化条件 $I_i \neq 0, i = 1, 2 \cdots, n$ 成立, 这一假定不仅是必要的, 而且是合理的, 数学定理在一些非退化条件下为真称为该定理一般性成立. 在整序过程中自动产生若干非退化条件, 是吴氏机械化原理的一项重要发现.

让我们从两个具体例子领略一下依据吴氏机械化原理实行数学定理机械化证明的大致过程.

例 1 （Desargues 定理）如图 1，设平面上二直线 l_1, l_2 相交于 O. l_1 上任取二点 A_1, A_2. l_2 上任取一点 B_1，过 A_2 作直线平行于 A_1B_1 且与 l_2 交于 B_2. 在平面上任取一点 C_1，过 A_2, B_2 分别作直线平行于 A_1C_1, B_1C_1，且二直线相交于 C_2，则 O, C_1, C_2 三点共线.

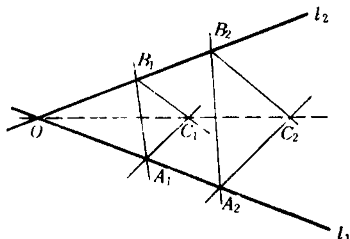


图 1

由于这一定理只涉及点线的从属关系及平行关系，故可如下选取坐标系：l_1 为一坐标轴，l_2 为另一坐标轴，点 O 为坐标原点，诸点坐标分别为

$$O(0,0), A_1(x_1,0), A_2(x_2,0), B_1(0,x_3),$$
$$C_1(x_4,x_5), B_2(0,x_6), C_2(x_7,x_8).$$

则定理的假设条件化为三个代数方程 $H_1=0, H_2=0, H_3=0$，其中

$$H_1 := x_1x_6 - x_2x_3,$$
$$H_2 := x_4(x_8 - x_6) - x_7(x_5 - x_3),$$
$$H_3 := (x_4 - x_1)x_8 - x_5(x_7 - x_2).$$

而结论可表示为 $G=0$，

$$G := x_4x_8 - x_5x_7.$$

依整序原理对多项式组 $\{H_1, H_2, H_3\}$ 进行整序，得到如下特征列：

$$A_1 := x_1x_6 - x_2x_3,$$
$$A_2 := (x_1x_3 - x_1x_5 - x_3x_4)x_7 + (x_4 - x_1)x_4x_6 + x_2x_4x_5,$$
$$A_3 := x_4x_8 - (x_5 - x_3)x_7 - x_4x_6,$$

其中初式为 $I_1 = x_1, I_2 = x_1x_3 - x_1x_5 - x_3x_4, I_3 = x_4$. 经直接计算

知 G 对 A_1,A_2,A_3 的余式为零，所以有如下的结论：在非退化条件 $I_1\neq 0, I_2\neq 0, I_3\neq 0$ 之下，Desargues 定理成立. 这样就完成了 Desargues 定理的机械化证明.

让我们考察一下各退化条件的几何意义，条件 $I_1=x_1=0$ 表明点 A_1 与点 O 重合，此时过 A_2 与 A_1B_1 平行的直线将与 l_2 平行，故定理假设呈一种不合理的退化. 条件 $I_2=x_1x_3-x_1x_5-x_3x_4=0$ 表明点 C_1 位于直线 A_1B_1 之上，因此点 C_2 必然在 A_2B_2 之上，并且位置是不确定的. 因而在这种退化情形定理未必成立. 条件 $I_3=x_4=0$，则表明点 C_1 位于 l_2 上，因而 C_2 也必然位于 l_2 上，定理也呈现为一种退化.

例 2 （初等几何中著名的 Feubach 定理）三角形的九点圆与任一与三角形三边都相切的圆相切.

如图 2，设三角形 $\triangle ABC$ 三边上的中点分别为 M,N,P；I 为与三边都相切的圆的圆心. 建立如下的坐标系：AB 为一坐标轴，过 I 与 AB 垂直的直线为另一坐标轴，则诸点坐标可取为：

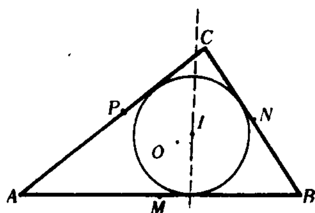


图 2

$I(0,x_1), A(x_2,0), B(x_3,0)$, $C(x_4,x_5)$，另设九点圆的圆心为 $O(x_6,x_7)$，九点圆的半径平方为 x_8^2，两圆心 O 与 I 的距离平方为 x_9^2. 则定理中诸假设条件化为

$$H_1:=x_9^2-x_6^2-(x_7-x_1)^2,$$
$$H_2:=(x_1^2-x_2^2)x_5+2x_1x_2(x_2-x_4),$$
$$H_3:=(x_1^2-x_3^2)x_5+2x_1x_3(x_3-x_4),$$
$$H_4:=4x_8^2-4x_7^2-(2x_6-x_2-x_3)^2,$$
$$H_5:=4x_8^2-(2x_6-x_2-x_4)^2-(2x_7-x_5)^2,$$
$$H_6:=4x_8^2-(2x_6-x_3-x_4)^2-(2x_7-x_5)^2,$$

而结论则表示为

$$G:=x_9^4+x_8^4+x_1^4-2x_9^2x_8^2-2x_9^2x_1^2-2x_8^2x_1^2.$$

经过对多项式组$\{H_1,H_2,\cdots,H_6\}$进行整序，可得特征列：

$$A_1:=(x_1^2+x_2x_3)x_4-x_1^2(x_2+x_3),$$

$$A_2:=(x_1^2+x_2x_3)x_5-2x_1x_2x_3,$$

$$A_3:=4x_6-(x_2+x_3+2x_4),$$

$$A_4:=4x_5x_7-x_5^2-(x_3-x_4)(4x_6-2x_2-x_3-x_4),$$

$$A_5:=4x_8-(2x_6-x_2-x_3)^2-4x_7^2,$$

$$A_6:=x_9^2-x_6^2-(x_7-x_1)^2.$$

直接计算可知G对$A_1,A_2,\cdots,A_6$的余式为零. 因此，依基本定理可知：在非退化条件$I_1=x_1^2+x_2x_3\neq 0$，$I_4=4x_5\neq 0$之下，Feubach定理成立. 大家知道，初等几何中Feubach定理的证明是相当难的，需要精巧的构思. 而依吴原理实行机械化证明，则是纯粹的代数运算，是简单易行的.

现在考察一下非退化条件的几何意义. 条件$I_1=x_1^2+x_2x_3=0$表明，三角形$\triangle ABI$为直角三角形. 即$\angle AIB$为直角，或者等价地有$\angle IAB+\angle IBA=90°$，然而直线$AI$，$BI$分别是$\angle CAB$和$\angle CBA$的分角线. 所以如果$x_1^2+x_2x_3=0$可推出$\angle CAB+\angle CBA=180°$，这显然表示三角形$ABC$处于一种退化状态；条件$I_4=4x_5=0$则表明点$C$位于第一坐标轴$AB$上，同样表明$\triangle ABC$处于一种退化状态. 因此假定$I_1\neq 0$，$I_4\neq 0$是合理的，必要的.

数学定理机械化证明的基本定理，对多项式组进行整序以获得特征列的整序定理以及零点结构定理是吴氏机械化原理的主要组成部分. 以这一原理为基础，在广泛的领域开展了机械化数学的研究. 依据这一原理，大量颇不简单的数学定理实现了机械化证明（参阅[9]，[10]）；一些难以证实的猜测得到证明，从而机械化地发现一些新的定理；实现了非欧几何定理的机械化证明、三角函数、超越函数恒等式的证明，不等式的证明以及微分几何定理的证明；开展了多项式理论的研究（参阅[11]－[16]）.

高次代数联立方程组的求解，是长期困扰数学家们的难题. 吴原理提供了高次联立方程组求解的有效方法(参阅[3],[4]). 对多项式组进行整序，本质上是一种消元过程，零点结构定理则保证求解过程中代数方程的解既不增加也不漏掉. 这一方法已经在非线性规划和线性控制系统的理论研究中获得成功的应用. 随着更多的人了解和掌握这一方法，必将在更为广泛的研究领域发挥作用.

吴原理为未知关系的推导提供了强有力的工具. 根据这一原理，一些著名公式实现了自动推导；某些运动轨迹得以自动确定；实现了多项式动力系统极限环的机械化计算，以及从 Kepler 定律到 Newton 定律的机械化推导.

吴氏原理自身也在迅速发展. 最为重要的是，吴文俊先生已经把代数多项式的机械化原理扩展到微分多项式，创立了代数微分几何的机械化原理. 这是机械化数学研究的重大进展. 微分多项式的机械化原理，必将对许多数学分支(诸如代数几何，微分几何，代数微分几何，微分方程等)的机械化研究产生深远的影响(参见[17],[18]).

代数几何是吴文俊机械化原理的数学基础. 吴氏原理着眼于多项式组的零点集，脱离了对多项式理想论的依赖，这样更能准确地反映问题的本质，准确有效地解决问题. 吴氏原理需要构造性代数几何的支持，反过来它又使代数几何中一些基本概念的建立更为直观、简捷，从而使得代数几何中许多问题的讨论得到简化. 存在性数学和构造性数学是数学发展中的两股潮流，它们相互交织，相辅相成，互相促进. 只强调其一而忽略其二，势必不利于数学的发展，那种片面追求数学抽象化形式化的倾向，必将把数学研究和从事数学研究的数学家们禁锢于象牙之塔，令数学失去活力. 实现数学的机械化将促使机械化数学的蓬勃发展. 数学机械化的思想将启示人们以更为开阔的视野去审视数学的现状，从新的角度来思考和勾画数学发展的前景.

当前，数学界正面对着非线性数学的强硬挑战. 物理学、力学、化学、生物科学、系统科学等领域涌现出大量非线性的数学问题，期待着数学家们予以解答. 数学理论本身，在许多分支中，诸如在代数几何、微分几何、微分方程、泛函分析、数学物理以及运筹学、统计学等分支中，有许多非线性数学问题迫切要求人们加以研究. 数学中的线性问题，从理论到方法都已相当成熟，数学家们运用起来得心应手. 而对于非线性问题却往往感到茫然无措. 长期以来，数学家们形成这样一种共识：线性问题和非线性问题是两个截然不同的领域，可以预想，非线性数学问题研究上的突破，必定要求数学家们对数学有更为深刻的认识和理解，从新的观念、新的高度来提出概念，提供方法，建立理论. 吴文俊机械化原理的创立为非线性数学问题的研究开辟了新的途径. 作为一种数学理论，吴原理本身就是研究和讨论非线性的数学问题. 随着吴原理的发展和完善，将为非线性数学问题的研讨和发展开创全新的局面.

所谓数学研究，或者是定理求证，或者是方程求解，或者是未知关系的推导. 吴文俊机械化原理在这三个方面都将大有作为. 它将促使数学研究的面貌改观. 它自创立之日起，就受到国际数学界的高度重视. 随着人们对它的了解日益深入，随着吴原理的应用更为广泛，随着它的影响逐步扩展，这一原理越来越受到国际数学界的高度评价. 吴文俊机械化原理植根于中国传统数学这片沃土，从《九章算术》、《四元玉鉴》到《机证原理》，中国传统的机械化数学思想清晰可见. 吴氏原理也注意汲取他人的有益成果，丰富自己. 任何庞大数学体系的建立，总是从最基本、最简单的问题入手的. 现在，初等数学的机械化问题已成为被翻越的一章. 更加诱人的前景也已展现在面前. 然而，数学机械化的道路毕竟是漫长而艰难的. 吴文俊机械化原理为数学的机械化奠定了坚实的基础. 建立具有中国特色的机械化数学体系仍然任重而道远. 这有待于有志于此的人们共同努力.

参 考 文 献

1. 吴文俊. 初等几何判定问题与机械化证明. 中国科学,1977:507～516.
2. 吴文俊. 几何定理机器证明的基本原理,科学出版社,1984.
3. 吴文俊.〈解方程器〉和〈Solver〉软件系统概述. 数学的实践与认识,1986(2).
4. 吴文俊.〈解方程器〉和〈Solver〉软件系统应用举例. 数学的实践与认识,1986(3).
5. 吴文俊编.《九章算术》与刘徽. 北京:北京师范大学出版社,1982.
6. Hilbert,D.. Grundlagen der Geometrie. 1899. Teuber.(中译本:几何基础. 科学出版社,1958.)
7. Tarski,A.. A Decision Methode for Elementary Algebra and Geometry. Berkeley. 1948～1951.(中译本:初等几何和代数的判定方法. 科学出版社,1959.)
8. 李约瑟. 中国科学技术史(第三卷). 科学出版社,1978.
9. 周咸青. Mechanical geometry theorem-proving. Reidel. 1988.
10. 王东明,高小山. Geometry theorems proved mechanically using Wu's method. 数学机械化与机械化数学研究报告,1987(2).
11. 王东明. On Wu's method for proving constructive geometric theorems. 数学机械化与机械化数学研究报告,1989(3).
12. 高小山. Transcendental functions and mechanical theorem proving in elementary geometries. 数学机械化与机械化数学研究报告,1987(2).
13. 吴文俊,吴天骄. A mechanization method of geometry and its application,5 ,Solving transcendental equations by algebraic method. 数学机械化与机械化数学研究报告,1989(3).
14. 李子明. A new proof of Collins's theorem. 数学机械化与机械化数学研究报告,1987 (1).
15. 刘卓军. A method of isolating complex roots of polynomials —— An application of Wu-Ritt Principle. 数学机械化与机械化数学研究报告,

1987(2).

16. 吴文俊. On reducibility problem in mechanical theorem proving of elementary. 数学月刊,1987(2). 数学机械化与机械化数学研究报告,1987(2).

17. 吴文俊. A mechanization method of geometry and its applications,4. Some theorem in planar kinematics. 系统科学与数学(英文版),1989(2).

18. 吴文俊. On the fundation of algebraic differential geometry. 数学机械化与机械化数学研究报告,1989(3).

堵丁柱
（中国科学院应用数学研究所）
Du Dingzhu
(Institute of Applied Mathematics,
Academia Sinica)

谈谈 P-NP 问题
A Tale on P-NP Problems

中国数学发展的若干主攻方向
Some Main Aspects on Development
of Mathematics in China

§1 什么是好算法

萝卜快了不去泥. 多了，滞销，就要遭到挑挑拣拣. 类似的情况发生在组合优化的研究中. 当算法的种类和数量积累到一定程度时，比较和分类的想法也自然产生了. 什么是好算法?J. Edmonds 于 1965 年提出了一个标准，它一直被沿用到今天. 那就是，计算时间作为输入数据长度的函数有一个多项式作为上界的算法，俗称多项式时间算法. 以计算时间作为衡量算法好坏的尺度是件较自然的事. 事实上，计算费用是和计算时间直接挂钩的. 但是，为什么用多项式而不用其他函数作分界点呢?这有两个原因.

其一是，常见的组合算法大致可分两类. 一类是多项式时间内可实现的；另一类需要指数时间，换句话说，对某些输入数据，它的计算时间是输入数据长度的指数函数. 随着自变量的增加，指数函数要比多项式的值增长的速度快得多. 因此，多项式时间算法的可实现性远大于指数时间算法. 这明显的区别是选择多项式作为好坏分界点的原因之一.

原因之二是，与计算模型的无关性. 计算时间总是相对于一种计算模型而言的. 读者可以把计算模型理解为一台理想的计算机. 由于理论上的需要，我们假定机器相当可靠，计算多久都不会出差错，同时假定机器内存空间总是足够大，不至于因空间不够而实现不了某算法. 常识告诉我们，对同一个算法来说，在不同类型的计算机上实现，计算时间是不同的. 但是，它们之间并非没有关系. 这种关系被称为广义 Chuch-Turing 命题. 它的内容如下：给定两计算模型 A 和 B，那么，存在一个多项式 P 使得对任何算法 α 有

$$A\text{ 实现 }\alpha\text{ 的时间} \leqslant p(B\text{ 实现 }\alpha\text{ 的时间}).$$

毫无疑问，好算法的概念应该与模型无关才适当. 由于多项式与多项式的复合函数是多项式，因此，模型 B 的多项式时间算法也

是模型A的多项式时间算法. 换言之,选多项式时间算法作为好算法的同义语恰好满足了与模型无关的要求.

当好算法的概念确定了以后,第一个令人关心的问题就是:是否每个组合优化问题都有好算法?在考虑优化问题的计算复杂性时,人们通常是把它化为相应的判定问题来考虑的. 判定问题是只有肯定和否定两种答案的问题. 例如,给出一个图,问它是否有Hamilton圈?这是个判定问题. 俗称Hamilton圈问题. 但是,组合优化问题通常不是判定问题. 例如,给出一个边长为自然数的图,问可以找到多短的旅行路线,从某顶点出发走遍所有顶点再回到出发点?这一优化问题被称为旅行售货员问题. 它的可能答案显然并非两种. 虽然优化问题不是判定问题,但是,很容易把它化为判定问题来处理. 例如,旅行售货员问题可以化为如下的形式:给出一个边长为自然数的图和一个自然数k,问是否有长度不超过k走遍所有顶点的巡回旅行路线?优化问题和相应之判定问题的难度是一样的. 换言之,一个有好算法,另一个也必有. 以旅行售货员问题为例,若原问题有好算法,以其解确定相应判定问题之解,可得相应判定问题之好算法. 若相应判定问题有好算法,那么原问题可以用如下方法求解:首先估计一个那种巡回行走路线的最短长度所在范围,例如说,它不会短于图的最小支撑树长a,也不会长于最小支撑树的两倍长$b=2a$,取$k=\left[\frac{a+b}{2}\right]$,解相应判定问题. 若答案为肯定的,则可判定最短巡回旅行路线长在区间$\left[a,\left[\frac{a+b}{2}\right]\right]$之中;若答案为否定的,则可判定最短巡回旅行路线长在区间$\left[\left[\frac{a+b}{2}\right],b\right]$之中,这样,用二分法,至多解$[\log_2(b-a)]$次相应判定问题,我们就可以找到原问题的解了. 显然,上述解法是个好算法. 由于二者之间的上述等价关系,在本文中我们不再区分优化问题与它的相应判定问题.

具有多项式时间算法的判定问题形成一个计算复杂性类，记为P. J. Edmonds 在提出算法好坏的衡量标准的同时，也注意到了有些组合优化问题可能不属于P，具体来说，他猜测，旅行售货员问题没有多项式时间算法.

§2 多项式时间归约

通常，人们把属于P的问题看作是易解问题，不属于P的问题看作是难问题. Edmonds 猜测又可以表达为，旅行售货员问题是难问题. 特别是，它比P中问题都难. 难易是就比较而言的，无比较何以谈难易呢？如果我们能建立两个问题之间的难易比较方法，无疑会对解决 Edmonds 猜测有所帮助. 至今为止，人们已经研究了许许多多种比较方法. 我们这里只介绍常见的一种，称为多项式时间归约，它起源于如下的考虑.

把一个问题通过演化归结为另一个问题来解是数学研究中的常用方法. 现在，我们看一下，如果可以把一个判定问题 A 化归为另一个判定问题 B，那么会怎样？这时，A 的答案可以通过如下途径找到：先把 A 化为 B，对 B 给出解答，其答案相应于 A 的解答. 因此，假如将 A 化为 B 费时不多，亦即在多项式时间内，那么 B 有多项式时间算法时，A 也一定有. 反之，若 A 没有多项式时间算法，则 B 也一定没有. 这确实从某种程度上反映了 A 和 B 之间的难易关系. 当问题 A 可以在多项式时间内化为问题 B 的特殊情况时，我们称 A 可多项式时间归约于 B，并且记为 $A \leqslant_m^p B$. 从上面的论述中易知

命题 1 若 $B \in \mathrm{P}$ 且 $A \leqslant_m^p B$，则 $\in \mathrm{P}$.

由于多项式与多项式的复合函数还是多项式，我们还有

命题 2 若 $A \leqslant_m^p B$ 且 $B \leqslant_m^p C$，则 $A \leqslant_m^p C$.

命题1说明多项式时间归约对类P是封闭的，命题2说明它是

个偏序关系. 为使读者对此归约有个较具体的印象，我们举两个例子.

例 1 Hamilton 圈问题 $\leqslant_m^p$ 旅行售货员问题.

给一个图 $G=(V,E)$，其中 V 是 G 的顶点集，而 E 是 G 的边集. 对 G 的每条边赋予单位长度，那么，当 G 有 Hamilton 圈时，在 G 上走遍全部顶点的最短巡回路线长为 $|V|$；当 G 没有 Hamilton 圈时，那样的旅行路线长至少要 $|V|+1$. 这样，对 G 来说，Hamilton 圈问题得到肯定的回答而且仅当对边长赋予 1 的 G 和 $k=|V|$，旅行售货员问题获肯定回答，亦即我们把 Hamilton 圈问题化成了旅行售货员问题. 另一方面，Hamilton 圈问题的输入是 G，由 G 构造边长均为 1 的 G 和 $k=|V|$（旅行售货员问题的输入）显然不需要多少时间，至多是 $|E|+|V|$ 的线性函数，即在多项式时间内，因此，Hamilton 圈 $\leqslant_m^p$ 旅行售货员问题.

由本例可体会到，所谓将问题 A 化为问题 B 的特殊情况，也就是将问题 A 的对象（输入）化为问题 B 的对象（输入）. 换句话说，找一个映象 f，从问题 A 的对象集合映入问题 B 的对象集合，使得问题 A 对对象 x 的答案是肯定的当且仅当问题 B 对 $f(x)$ 获肯定答案. 当 f 是多项式时间可算时，我们就有 $A\leqslant_m^p B$ 了.

在接触第二个例子之前，首先让我们介绍一个问题.

有三颗珍珠，其中可能有假的. 某人想选一颗真的. 他征求了三位识货者的意见. 将珍珠编号后，甲说："1 号和 2 号是真的."乙说："2 号和 3 号是真的."丙说："3 号为真而 2 号为假."假如每人都至少说对了一半，问：该人应选哪颗珍珠？

这问题可用逻辑方程求解. 令布尔变量 x_1，x_2 和 x_3 分别代表三颗珍珠，其意义如下：

$$x_i=\begin{cases}1，若\ i\ 号珍珠为真，\\0，若\ i\ 号珍珠为假.\end{cases}$$

按题意，每人的意见都至少对一半，故有

$$x_1 + x_2 = x_2 + x_3 = x_3 + \overline{x}_2 = 1. \tag{2.1}$$

这里的加法是布尔加法运算，即 $0+1=1+0=1+1=1$ 且 $0+0=0$. $\overline{x}_2$ 表示变量 x_2 的非，即 $x_2=0$ 时 $\overline{x}_2=1$ 并且 $x_2=1$ 时 $\overline{x}_2=0$. (2.1) 式也可写为如下形式：

$$(x_1 + x_2)(x_2 + x_3)(x_3 + \overline{x}_2) = 1. \tag{2.2}$$

将其展开，可得

$$x_1x_2x_3 + x_1x_3 + x_1x_3\overline{x}_2 + x_2x_3 = 1.$$

这说明，方程(2.2) 的每组解均含 $x_3 = 1$，因此，该人应选 3 号珍珠.

如同方程(2.2) 左端形式的布尔表达式被称为合取范式. 合取范式是布尔变量和布尔变量非的布尔和之积. 在合取范式中，每个因子都被称为初等和. 当一个布尔函数有使其为 1 的赋值时，被称为可满足的. 给出一个合取范式，问它是否可满足?这一问题称为满足问题.

例 2 满足问题 $\leqslant_m^p$ Hamilton 圈问题.

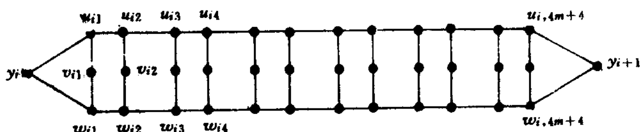


图 1 H_i

设 $F = C_1 \cdot C_2 \cdot \cdots \cdot C_m$ 是个含 n 个布尔变量 $x_1, x_2, \cdots, x_n$ 的合取范式. 我们要由 F 出发构造一个图 $G(F)$，使得 F 可满足当且仅当 $G(F)$ 有 Hamilton 圈. 首先，对每一变量 x_i，构造一如图 1 的梯图 H_i. 我们注意，该图恰有两条从 y_i 到 y_{i+1} 的 Hamilton 路，亦即经过 H_i 所有顶点的路，$P_{i1} = (y_i, u_{i1}, v_{i1}, w_{i1}, w_{i2}, v_{i2}, u_{i2}, \cdots, u_{i,4m+4}, y_{i+1})$ 和 $p_{i0} = (y_i, w_{i1}, v_{i1}, u_{i1}, u_{i2}, v_{i2}, w_{i2}, \cdots, w_{i,4m+4}, y_{i+1})$. 将诸 H_i 连结成一圈(如图 2)，形成图 H. H 有 2^n 个 Hamilton 圈恰

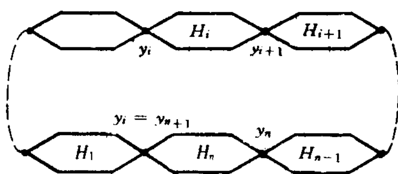


图 2 H

好相应于 n 个布尔变量 x_1, $x_2,\cdots,x_n$ 的 2^n 种不同的赋值. 当某 Hamilton 圈经过 p_{i0}(或 p_{i1})时,相应的赋值是 $x_i=0$(或 1). 现在,对每一初等和 c_j,给出一新点 z_j,并且,对 c_j 含的每个 x_i,将 z_j 与 $u_{i,4j-2}$ 和 $u_{i,4j-1}$ 相连;对 c_j 含的每个 $\overline{x}_i$,将 z_j 与 $w_{i,4j-2}$ 和 $w_{i,4j-1}$ 相连. 这样所得的含 $12mn+m+14n$ 个顶点的图,我们定义为 $G(F)$. 显然,由 F 构造 $G(F)$ 可以在关于 $m+n$ 的多项式时间内完成.

设 F 是可满足的,并且设使 F 满足的赋值对应的图 H 的 Hamilton 圈为 Q. 我们来修改 Q,使其成为 $G(F)$ 的 Hamilton 圈. 注意,每个 c_j 必含一项被赋于值 1,若此项为 x_i,则 Q 必含 p_{i1},而且 $G(F)$ 必含边 $(u_{i,4j-2},z_j)$ 和 $(u_{i,4j-1},z_j)$. 我们将此二边添入 Q 且去掉边 $(u_{i,4j-2},u_{i,4_{j-1}})$,可使 Q 多经过一顶点 z_j. 若 c_j 含 $\overline{x}_i$ 且 $\overline{x}_i=1$,则 Q 必经过 p_{i0},而且 $G(F)$ 必含边 $(w_{i,4j-1},z_j)$ 和 $(w_{i,4j-1},z_j)$. 将此二边加入 Q 且去掉边 $(w_{i,4j-2},w_{i,4j-1})$,也可以使 Q 经过 z_j. 这样,就找到了 $G(F)$ 的 Hamilton 圈.

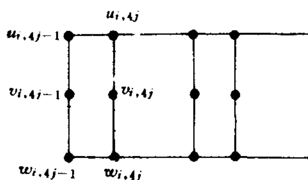


图 3

反之,设 $G(F)$ 有 Hamilton 圈 Q. 对每个 z_j,它在 Q 中只能与某个 $u_{i,4j-2}$,或者 $u_{i,4j-1}$,或者 $w_{i,4j-2}$,或者 $w_{i,4j-1}$ 相邻. 不妨设它与 $u_{i,4j-2}$ 相邻,亦即 Q 含边 $(u_{i,4j-2},z_j)$,那么 Q 必含边 $(u_{i,4j-2},v_{i,4j-2})$ 和 $(v_{i,4j-2},w_{i,4j-2})$,否则的话,Q 无法经过点 $v_{i,4j-2}$. 这样,假若 Q 不含边 $(u_{i,4j-1},z_j)$,那么 Q 要经过 $v_{i,4j-1}$ 和 $v_{i,4j}$ 都必须通过 $u_{i,4j}$ 和 $w_{i,4j}$ 进

出(如图3),这是不可能的,故Q必含边$(u_{i,4j-1},z_j)$. 这说明,z_j在Q中的两个相邻点之间有H中的边相连. 去掉z_j及其相邻边,添入连结z_j的相邻点的边,我们可由Q缩至H的一个Hamilton圈. 容易检验,该圈相应的赋值可使F满足.

§3 P-NP 问题

设$\mathscr{C}$是由判定问题组成的一个类. 若$B \in \mathscr{C}$且对任何$A \in \mathscr{C}$有$A \leqslant_m^p B$,那么称B为$\mathscr{C}$完全问题. 如果一个算法的计算时间,作为输入数据长度n的函数,以$2^{\alpha n}$为上界,其中α为一常数,那么称该算法是指数时间的. 所有具有指数时间算法的判定问题组成一个类,记为EXP. 1965年,Hartmani's和Stearns以Turing机为计算模型奠定了计算复杂性理论的严格理论基础. 随之,人们证明了$\mathrm{P} \subsetneq \mathrm{EXP}$. 旅行售货员问题属于EXP是个众所周知事实. 一个自然的想法是,如果能证明它是EXP完全的,那么依据命题1,它就不属于P了. 很可惜,这种想法至今还是难以实现的. 但是,1971年,Cook证明了一个类似的结果.

定理1 满足问题是NP完全的.

什么是NP类?NP类是具有不确定型多项式时间算法的判定问题组成的类. 所谓不确定型是种计算方式. 现实世界中的计算机的计算方式都是确定型的. 事实上,如果机器和程序没有毛病,那么只要输入一确定,计算就会按部就班地一步步走下去. 无论算几遍,计算过程都不会改变. 可是,不确定型计算可以不具备上述可重复的特点. 下面是解满足问题的不确定型算法. 我们来分析一下它的特点.

第1步. 输入合取范式$F(x_1,\cdots,x_n)$. 猜变量$x_1,\cdots,x_n$的一组赋值.

第2步. 检查F在该赋值下是否值为1. 若值为1,则停止,输出

1；否则，让计算进入无穷循环.

首先，该算法有一个猜的步骤.猜的结果不同，后面的计算过程也就不同了.这意味着计算过程不是唯一的，不具备可重复性.其次，若 F 可满足，则至少有一计算道路(猜对时)使计算在有限步内停止并且输出 1；若 F 不可满足，则在有限步内停止且输出 1 的计算过程是不存在的.

在估计不确定型算法的计算时间时，不失一般性，可以假定，从每两个元素中猜一个，要算一步.因此，上述算法实际上是做了 n 次猜测.不确定型算法的计算时间定义为各种可能的计算过程的最短时间.一个解判定问题的不确定型算法被称为多项式时间的，只要计算时间，作为定义在使问题得到肯定回答的输入上的函数，有一多项式为上界.当变量的赋值一旦猜定，计算合取范式 F 的值显然只花多项式时间，因此，上述算法是不确定型多项式时间算法.

一般来说，不确定型多项式时间算法正是有前述两特点的多项式时间算法.下面是又一个例子.

第 1 步.输入一图 G，猜全部顶点的一个顺序.

第 2 步.若 G 有按此顺序的一个圈，则计算停止且输出 1；否则，让计算进入无穷循环.

这是个解 Hamilton 圈问题的不确定型多项式时间算法.在估计猜占用的时间时，可以这样来想：将每个顶点对应一个二进位整数.设 G 有 n 个顶点，那么我们只要用长为 $[\log_2 n]$ 的二进制整数就够了.这样，每猜一个顶点，只要猜 $[\log_2 n]$ 个数字，即花 $[\log_2 n]$ 步.因此，猜全部顶点的顺序至多要 $n[\log_2 n]$ 步.

前面两个算法说明满足问题和 Hamilton 圈问题都属于 NP.作为一个练习，读者可以证明旅行售货员问题也属于 NP.根据定理 1 以及前一节的例 1 和例 2，可以进一步知道，Hamilton 圈问题和旅行售货员问题也是 NP 完全的.

自 Cook 定理发表以后，到今天，人们已发现了二千多个 NP 完全问题，这些问题分布于国民经济的各部门以及逻辑、代数、自动机、组合优化等许多领域. 按照 NP 完全性的定义和命题 1 可知，一个 NP 完全问题属于 P 当且仅当所有 NP 完全问题属于 P，而且，它们不属于 P 当且仅当 $P \neq NP$. 这一广泛的应用背景已使 P=?NP 问题成为当前的一个重大而著名的题目.

P=?NP 又称 P-NP 问题，这是个难题. 围绕这一难题，人们已做出了相当多的工作. 下面我们介绍比较活跃的两个方面.

[注] 本节和上一节的叙述没有完全按照历史的发展顺序. 事实上，Cook 定理出现在本文所述的多项式时间归约出现之前. 后者是 1972 年由 Karp 引入的，因此又称 Karp 归约. Cook 于 1971 证明的 NP 完全性是相对于另一种归约而言的. 它今天被称为多项式时间 Turing 归约或者 Cook 归约.

§4 多项式时间同构

在前两节中，我们证明了满足问题 $\leq_m^p$ Hamilton 圈问题，亦即存在一个多项式时间可计算的映射 $G(\cdot)$，将每个合取范式 F 映到一图 $G(F)$，使得 F 可满足当且仅当 $G(F)$ 有 Hamilton 圈. 同时，我们说明了满足问题和 Hamilton 圈问题都是 NP 完全问题. 作为一个推论，我们有 Hamilton 圈问题 $\leq_m^p$ 满足问题，亦即存在一个多项式时间可计算的映射 $F(\cdot)$，将每个图 G 映到一合取范式 $F(G)$，使 G 有 Hamilton 圈当且仅当 $F(G)$ 可满足. 一般来说，$G(\cdot)$ 和 $F(\cdot)$ 没有明显的关系. 但是，1975 年，Berman 和 Hartmanis 发现了，存在一对 $G(\cdot)$ 和 $F(\cdot)$，使 $G(\cdot)$ 是 $F(\cdot)$ 的逆函数. 设 A 和 B 是两个判定问题. 如果存在一个由 A 的对象集到 B 的对象集上的一一且满的映射 f，使得 f 和 f^{-1} 都是多项式时间可算的，并且针对问题 A，x 获肯定，当且仅当针对问题 B，$f(x)$ 获肯定，那么称 A 和 B 是

多项式时间同构的. 因此，我们也可以说，满足问题和 Hamilton 圈问题是多项式时间同构的. 事实上，Berman 和 Hartmanis 发现了，人们在实际中遇到的 NP 完全问题都是多项式时间同构的，同时做了如下猜测.

Berman-Hartmanis 猜测 全部 NP 完全问题都是多项式时间同构的.

为叙述方便，我们将简称它为 B-H 猜测. B-H 猜测可以导出 P $\neq$ NP. 要明白这一点，只要了解如下的不难证明的事实.

命题 3 若 P = NP，那么除去 $\varnothing$ 和 Σ^* 两个特例外，P 中问题都是 NP 完全的，其中 $\varnothing$ 是不可能有肯定回答的问题，而 Σ^* 是不可能有否定回答的问题.

现在，考虑这样的判定问题 A：给出一自然数 n，问它是否等于 1？这只有 $n = 1$ 时才会获得肯定的回答. 很明显，$A \in$ P. 假若 P = NP，依命题 3，A 是 NP 完全的. 这样，B-H 猜测可导出 A 和满足问题多项式时间同构. 这意味着在一个有限集 $\{1\}$ 和无穷集 $\{F \mid F$ 是可满足的合取范式$\}$ 之间存在一一对应关系，这显然不可能，因此，B-H 猜测导出 P $\neq$ NP. 正由于此，B-H 猜测引起了许多人的研究兴趣. 1980 年，Mahaney 证明了 B-H 猜测的一个重要推论.

定理 2(Mahaney) 若 P $\neq$ NP，则 NP 完全问题均非稀疏的.

什么是稀疏的？简单地说，就是获肯定回答的可能性很小的问题. 为了给出较精确的说法，我们首先说明输入数据长度一词的精确含义. 计算机只能直接存储 0 和 1 两种符号. 其余的输入对象，如合取范式，图等，都是先编码为由 0 和 1 组成的有限符号串，然后存到存储器中的. 所谓输入数据长度就是上述符号串的符号个数，或者说在存储器中占有的位置个数. 长度不超过 n 的 0-1 符号串有 $2^{n+1} - 1$ 个. 如果其中只有不超过 $p(n)$ 个符号串对应获肯定回答之输入，而且 $p(n)$ 是个多项式，那么与 $2^{n+1} - 1$ 相比较，获肯定回答的可能性是太少了，因此，称那样的问题是稀疏的. 人们已

知

命题 4　若 A 和 B 是多项式时间同构的. 则 A 稀疏当且仅当 B 稀疏.

命题 5　满足问题非稀疏.

由此二命题可推知

推论　若 B-H 猜测真,则所有 NP 完全问题是非稀疏的.

Mahaney 定理是 B-H 猜测的一个弱推论. 可是,它是人们已知的 NP 完全问题的不多性质之一. Mahaney 定理加强了人们对 B-H 猜测的信任,但是,1983 年,Joseph 和 Young 提出了相反的见解. 假如把每个判定问题的输入数据都编码为 0-1 符号串,那么所有获肯定回答的输入就组成了一个 0-1 符号串集. 这一集反过来也可以确定该判定问题. 因此,在理论上,我们可以用这些 0-1 符号串集来代替判定问题来讨论. 相应于 NP 完全问题的 0-1 符号串集被称为 NP 完全集. Joseph 和 Young 构造了一种新的 NP 完全集,称为 k- 创造集. 种种迹象表明,k- 创造集之间多项式时间同构的可能性与单陷门函数的存在性有关连. 什么是单陷门函数?设 f 是一个定义在由全部有限长 0-1 符号串组成的集 Σ^* 上且值域也在 Σ^* 中的一对一的函数. 如果 f 满足如下条件

(1) 存在多项式 p 使 $|x| \leqslant p(|f(x)|)$,其中 $|x|$ 和 $|f(x)|$ 分别表示符号串 x 和 $f(x)$ 的长度,

(2) f 是多项式时间可计算的,但是,f^{-1} 不是多项式时间可计算的,

那么称 f 是单陷门函数. 单陷门函数的存在性是理论密码学的重大未解问题. 人们已经知道

命题 4　若单陷门函数存在,则 $p \neq \mathrm{NP}$.

Joseph 和 Young 猜测,B-H 猜测不真当且仅当单陷门函数存在. 这样就构成了如图 4 的有趣关系.

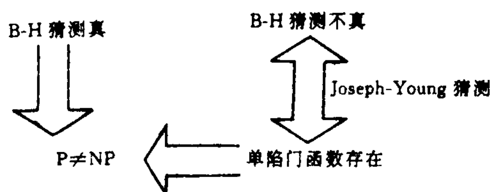


图 4

由图 4 中关系可知，Joseph-Young 猜测也可以导出 P ≠ NP. 1985 年，Ko，Long 和 Du 证明了单陷门函数的存在性确实与多项式时间同构有关，从而支持了 Joseph-Young 猜测. 但是，无论是 B-H 猜测，还是 Joseph-Young 猜测，它们的最后命运还等待着新的研究结果来决定.

§5 电路复杂性

我们已经谈到过，每个判定问题对应于一个 0-1 符号串集. 现在，我们指出，每一个 0-1 符号串集 A 又可对应于一个布尔函数序列 $\{f_n\}$，其中 f_n 是 n 元函数，定义如下：

$$f_n(x_1,\cdots,x_n)=\begin{cases}1, 若\ x_1x_2\cdots x_n\in A,\\0, 若\ x_1x_2\cdots x_n\notin A,\end{cases}$$

每个布尔函数 f_n 都可以用由与门、或门和非门组成的逻辑电路来计算（如图 5）. 这里，我们规定每个逻辑门的输入端线不超过两根. 这时，逻

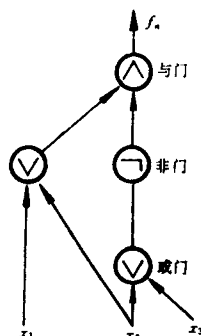


图 5

$f_n=(x_1+x_2)\cdot(\overline{x_2+x_3})$

辑电路的尺寸定义为逻辑门的个数，并且布尔函数 f_n 的电路复杂性定义为计算它的电路的最小尺寸 $s(n)$. 如果存在一个多项式，使得对任何 n 有 $s(n) \leqslant p(n)$，那么称集合 A（或相应判定问题）有多项式尺寸的电路. 如下的事实是相当重要的.

定理 3 P 中问题都有多项式尺寸电路.

定理 4 如果存在一个 NP 完全问题有多项式尺寸电路，那么 $\triangle_2^p = \text{HP}$.

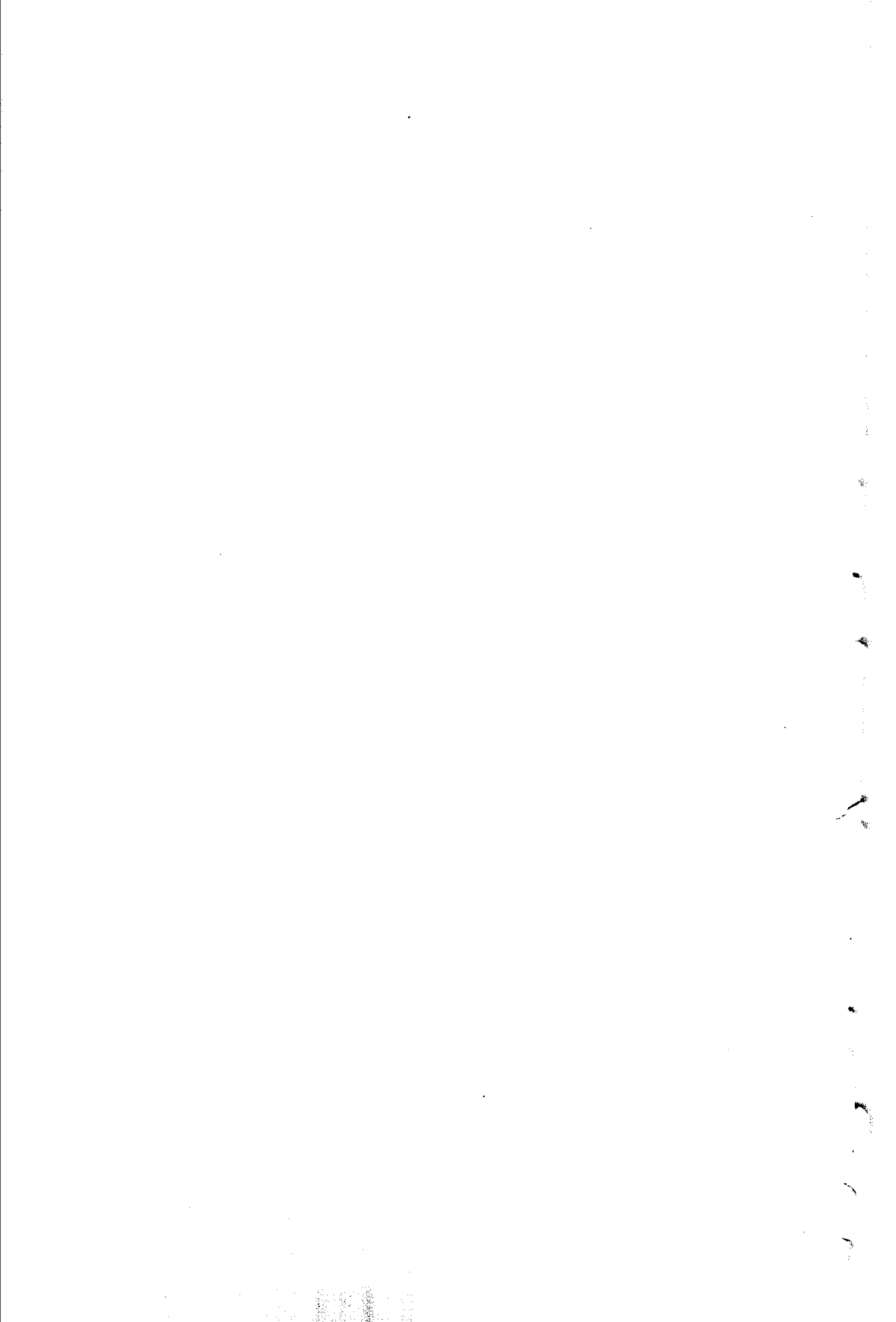
这里，$\triangle_2^p$ 和 HP 都是计算复杂性类. 如同 P-NP 问题一样，$\triangle_2^p = ?\text{HP}$ 也是个未解问题，但是，人们普遍认为 $\triangle_2^p \neq \text{HP}$. 这样，NP 完全问题有多项式尺寸电路的希望是不大的. 这提供了研究 P-NP 问题的一个途径. 也就是说，通过证明某一 NP 完全问题没有多项式尺寸电路达到证明 $\text{P} \neq \text{NP}$ 的目的. 1985 年，Razhorove 考虑图的最大团问题，这是个 NP 完全问题. 他证明了，最大团问题没有多项式尺寸的不用非门的电路（这种电路称为单调电路）. 令人惋惜的是，Razhorove 的方法不能应用到非单调电路上. 事实上，即使只允许使用一个非门，人们至今也证明不了相应的结果. 如果读者有兴趣，不妨来试试，新的突破正在等待着你们.

参 考 文 献

1. R. V. Book, Studies in Complexity Theory, John Wiley & Sons, Inc., New York, 1986.
2. U. Schöning, Complexity and Structure, Springer-Verlag, New York, 1985.
3. K. Mehlhorn, Graph Algorithm and NP-Completeness, Springer-Verlag, New York, 1984.

Ⅱ

非线性分析
Non-Linear Analysis



齐民友
(武汉大学数学系)
Qi Minyou
(Department of Mathematics,
Wuhan University)
陈恕行
(复旦大学数学研究所)
Chen Shuxing
(Institute of Mathematics,
Fudan University)

微局部分析简介
Introduction to Microlocal Analysis

中国数学发展的若干主攻方向
Some Main Aspects on Development
of Mathematics in China

§1 从微局部观点考察函数

与物理学中有宏观物理学与微观物理学相仿，在数学理论的分析学中也有大范围分析与微局部分析等分支.

函数是分析学中所研究的一类基本对象，在函数的各种性质（例如相等、可微性、有界性、振⊢性、渐近性等）中，有些性质是局部性质，例如我们可以说一个函数在某点可以求导；有些性质是整体性质，例如我们可以说一个函数在某区域中有界. 而函数的某些整体性质常常可以从其每个局部中的局部性质推出，例如，若两个函数在某区域中点点相等，则这两个函数在该区域中相等. 近代数学理论中，函数概念已发展成广义函数（分布）或更一般的概念. 这时，局部性质就是指在一点的充分小的邻域中的性质，这种局部性质也常常可推出相应的整体性质. 以下为了叙述的方便，我们所提及的函数一般就是指广义函数，而说到函数在某点的性质就是指它在该点充分小的邻域中的性质.

可微性也是一种局部性质，为考虑函数在某定义域中的可微性，可以分别考察它在各点的可微性. 如果一函数 u 在某点为无穷次可微的，我们称它为在该点是 C^∞ 光滑的，或简称为在该点光滑，否则，则称它在该点非光滑. 所有非光滑点全体称为函数的奇支集，记为 sing supp u. 进一步的考察可知，在一点附近同为非光滑的函数，其性质可以相差很大. 这种现象促使我们对函数的局部性质作更细致的考察.

我们以 x_1, x_2 平面上的阶梯函数 $\theta(x)$ 与 Dirac 函数 $\delta(x)$ 为例进行讨论. $\theta(x)$ 定义为

$$\theta(x) = \begin{cases} 1, & x_1 > 0; \\ 0, & x_1 \leqslant 0. \end{cases} \tag{1}$$

它在 $x_1 = 0$ 坐标轴上有一个台阶，$\delta(x)$ 是一个广义函数，它可以

看成一列函数

$$\delta_n(x) = \begin{cases} \dfrac{n^2}{\pi}, & |x| < \dfrac{1}{n}; \\ 0 & |x| \geqslant \dfrac{1}{n} \end{cases} \tag{2}$$

的弱极限，从而可以想象成在原点有一个高度为无限大的山峰，其余点为零. 显然，这两个函数在原点附近的局部性质相去甚远. 特别，当我们把这两个函数取为波动方程的初值时，所得到的波动方程的解是截然不同的.

初值问题

$$\begin{cases} u_{tt} = u_{x_1x_1} + u_{x_2x_2}; \\ u|_{t=0} = \theta(x); \\ u_t|_{t=0} = 0 \end{cases} \tag{3}$$

的解是

$$u = \begin{cases} 1, & x_1 > t; \\ \dfrac{1}{2}, & -t < x_1 < t; \\ 0 & x_1 < -t. \end{cases} \tag{4}$$

而初值问题

$$\begin{cases} v_{tt} = v_{x_1x_1} + v_{x_2x_2}; \\ v|_{t=0} = \delta(x); \\ v_t|_{t=0} = 0 \end{cases} \tag{5}$$

的解是

$$v = \begin{cases} \dfrac{1}{2\pi} \dfrac{\partial}{\partial t} \dfrac{1}{\sqrt{t^2 - x^2}}, & |x| > t; \\ 0, & |x| < t. \end{cases} \tag{6}$$

由解的表达式可见，问题(3)的解 u 的奇性(非光滑点)沿着平面 $x_1 = \pm t$ 往两边传播，而问题(5)的解 v 的奇性从原点出发沿特征锥 $|x| = t$ 往各个方向传播.

怎样把 $\theta(x)$ 与 $\delta(x)$ 在原点非光滑性的不同特点表达清楚呢?一个很好的工具就是 Fourier 变换. Fourier 变换将函数的可微性与变换后的象在无穷远处的增长性联系在一起. 我们知道 Fourier 变换建立了速降函数空间 φ 到它自身的同构,故若 $u(x)$ 为 C_0^∞ 函数,则 $u(x)$ 的 Fourier 变换 $\hat{u}(\xi)$ 在 $|\xi| \to \infty$ 时是速降的. 从而对任意 N,有常数 C_N,使得 $|\hat{u}(\xi)| \leqslant C_N(1 + |\xi|)^{-N}$ 成立,或证为

$$|\hat{u}| = 0(|\xi|^{-N}), \qquad |\xi| \to \infty. \tag{7}$$

于是,如果 $u(x)$ 在某点为 C^∞,我们就可以找一个支集在该点充分小的邻域中的 C_0^∞ 函数 $\varphi(x)$,使 $\varphi(x)u(x)$ 的 Fourier 变换 $\widehat{\varphi u}(\xi)$ 满足(7)式,反之,若 u 在一点不光滑,则对任意在该点非零的 C_0^∞ 函数 $\varphi(x)$, $\widehat{\varphi u}(\xi)$ 不可能有(7)中所示的估计.

但是,(7)式提供了我们作更细致考察的可能性,在(7)式不成立的情形下, $\hat{u}(\xi)$ 仍有可能沿着某些方向是速降的. 例如,该 $\varphi_1(x_1)$, $\varphi_2(x_2)$ 都是 C_0^∞ 函数,则 $\theta\varphi_1\varphi_2$ 的 Fourier 变换是

$$\begin{aligned} F[\theta\varphi_1\varphi_2] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-i(x_1\xi_1 + x_2\xi_2)}\theta(x_1\varphi_1(x_1)\varphi_2(x_2)\mathrm{d}x_1\mathrm{d}x_2 \\ &= \hat{\varphi}_2(\xi_2)\int_0^{\infty} e^{-ix_1\xi_1}\varphi(x_1)\mathrm{d}x_1. \end{aligned}$$

由于 $|\xi_2| \to \infty$ 时 $\hat{\varphi}_2(\xi_2)$ 为速降的,而后一个因子一般不是速降的,注意到 $\left|\int_0^{\infty} e^{-ix_1}\xi_1\varphi_1(x_1)\mathrm{d}x_1\right| \leqslant \int_0^{\infty} |\varphi_1(x_1)|\mathrm{d}x_1$ 是有界量. 所以,若记 $(\xi_1, \xi_2) = |\xi|(\tau_1, \tau_2)$,只要 $\tau_2 \neq 0$,就有 $|\xi| \leqslant c|\xi_2|$,故在 $|\xi| \to \infty$ 时仍有

$$F[\theta\varphi_1\varphi_2](\xi) = O(|\xi_2|^{-N}) = O(|\xi|^{-N}).$$

由此可知,仅在 $\tau_2 = 0$ 时, $\theta\varphi_1\varphi_2$ 的 Fourier 变换才不是速降的,从而我们可以说,只有方向 $(\tau_1, 0)$(或直接写成 $(\pm 1, 0)$)是"坏"方向,而其他方向均是"好"方向.

对于函数 $\delta(x)$,情形就完全不同了,只要 $\varphi(0) \neq 0$,就有

$$F[\varphi\delta](\xi) = \langle \varphi\delta, e^{-ix\xi} \rangle = \varphi(0) \neq 0.$$

因此，不管 ξ 沿什么方向趋于无穷，$\hat{\delta}(\xi)$ 都不会速降，也就是说，这时所有方向都是“坏”的. 与前面所给出的二维波动方程初值问题的解相比较，我们发现，“坏”的方向与解的奇性传播的方向有密切的联系. 对问题(3)，在 $x_1 = 0$ 上各点的奇性以波速 1 向方向 $(\pm 1, 0)$ 传播，因而波阵面为 $x = \pm t$. 对问题(5)，虽然初值的奇性仅在原点出现，但它以波速 1 向四周传播，从而解在特征锥 $|x| = t$ 上出现奇性. 这里所指出的波动方程初始值问题中初始值 Fourier 变换的“坏”方向与初始值奇性传播方向间的联系决非偶然，它正是微局部分析理论将揭示的一个规律. 为了精确地描述函数 u 在非光滑点的局部结构，Hörmander 引入了以下的波前集 WFu 的概念.

定义 1 对于给定的 $u \in \mathscr{D}'(\Omega)$，WF$u$ 为 $\Omega \times R^n_\xi \setminus O$ 中这样的子集：若 $(x_0, \xi_0) \notin \mathrm{WF}u$，则有 x_0 的邻域 ω，ξ_0 的锥邻域* V，使得对任一函数 $\varphi(x) \in C_0^\infty(\omega)$，任意 N，存在常数 C_N，有

$$|\widehat{\varphi u}(\xi)| \leqslant C_N(1 + |\xi|)^{-N}, \quad \forall \xi \in V \tag{8}$$

成立.

显然，若 $(x_0, \xi_0) \in \mathrm{WF}u$，则 $x_0 \in \mathrm{sing\,supp}\, u$.

又利用有限覆盖定理可证，如果对一切方向 ξ，$(x_0, \xi) \notin \mathrm{WF}u$，则 $x_0 \notin \mathrm{sing\,supp}\, u$，于是，若以 Π 表示从 $R^n_x \times R^n_\xi$ 到 R^n_ξ 的投影算子，我们有

$$\Pi\ \mathrm{WF}u = \mathrm{sing\,supp}\, u. \tag{9}$$

波前集 WFu 又称奇谱，它反映了函数 u 的一个重要特点，引入这一概念也有助于研究函数的运算. 例如，广义函数的乘法是一个很困难的问题，而今利用波前集的概念，可以得到广义函数能够

---

* 所谓锥邻域是指这个邻域为锥形的，即若 $\xi \in V$，则对任意 $t > 0$，也有 $t\xi \in V$.

作乘积的一个充分条件. 它就是:若 $u,v \in \mathscr{D}'(\Omega)$ 且

$$\begin{aligned} &\mathrm{WF}u + \mathrm{WF}v \\ &= \{(x,\xi);\xi = \xi_1 + \xi_2,(x,\xi_1) \in \mathrm{WF}u,(x,\xi_2) \in \mathrm{WF}v\} \end{aligned}$$

非空,则乘积 uv 可定义.

对函数作为对象进行各种运算所导致的相应波前集的变化规则已有较充分的研究,例如可参见[1]. 特别,当我们作自变数的可逆 C^∞ 变换 $x \xrightarrow{\varphi} y$ 时,$u(y)$ 与$(u\circ\varphi)(x)$ 波前集之间有关系

$$\mathrm{WF}(u\circ\varphi) = \{(x,\xi);x = \varphi^{1}(y),\xi = {}^t\varphi'(x)\eta,(y,\eta) = \mathrm{WF}u\}. \tag{10}$$

等式(10) 中所示的(x,ξ) 与(y,η) 之间的关系恰好是余切丛的局部坐标的变换规律,因此我们也可对给定在一个微分流形 M 上的函数来定义波前集,它是余切丛 T^*M 上的一个子集.

我们通常把自变数 x 变化的区域 Ω 或流形 M 称为底空间,而把变量 ξ 变化的空间 R^n_ξ 称为对偶空间或纤维空间. 在前面的讨论中,我们是先在底空间中对所讨论的函数作局部化,再对对偶空间中所讨论对象作进一步的局部化处理. 两者合之称为微局部的处理. 上述过程有时也可以逆过来进行,例如,设 $u(x)$ 为给定的函数,作其 Fourier 变换 $\hat{u}(\xi)$(若设 $u(x)$ 具有紧支集,则 Fourier 变换必定存在),又在 R^n_ξ 中找一个单位分解 $1 = \sum\varphi_j(\xi)$,其中 $\varphi_j(\xi) \in C_0^\infty(R^n_\xi)$,则可以有分解 $u(x) = \sum u_j(x)$,其中

$$\hat{u}_j(\xi) = \varphi_j(\xi)\hat{u}(\xi).$$

$u(x)$ 的许多性质可以通对每个 $u_j(x)$(或 $\hat{u}_j(\xi)$) 的了解而得到,而 $u_j(x)$ 的性质又依赖于对偶空间中局部化(即单位分解) 的方式,当然,若有必要的话,还可以在 R^n_x 中对 $u_j(x)$ 再作分解.

同时在底空间上与对偶空间上考察函数是一个观点上的飞跃. 它相当于物理学中不仅从位置的变化而且也从动量的变化来研究物体的运动. 这是量子物理学的基本思想之一. 与此相应,在

数学中也产生了既在底空间中又在对偶空间中作局部化处理的研究方法，这种方法就称为微局部分析方法.

§2 拟微分算子

若要对给定的函数同时在底空间与对偶空间中进行微局部化，就需要引入新的数学工具，其中最基本的工具就是拟微分算子. 拟微分算子是微分算子概念的推广，它的一个特点也正是把底空间与其对偶空间作完全平等的对待. 其系统理论是于 60 年代中期建立起来的.

在 R^n 中偏微分算子的一般形式为 $P = \sum\limits_{|\alpha|\leqslant m} a_\alpha(x)D^\alpha$，这里 α 为重指数$(\alpha_1,\cdots,\alpha_n)$，$D^\alpha = D_{x_1}^{\alpha_1}\cdots D_{x_n}^{\alpha_n}$，且为了以后运算的方便我们采用了微分记号 $D_{x_j} = \dfrac{1}{i}\dfrac{\partial}{\partial x_j}$. 形式上看，$P$ 是 $D_{x_1},\cdots,D_{x_n}$ 的一个多项式，由 Fourier 变换的性质知

$$Pu = \sum_{|\alpha|\leqslant m} a_\alpha(x)D^\alpha u = \int e^{ix\cdot\xi}p(x,\xi)\,\hat{u}\,(\xi)\,đ\xi, \tag{11}$$

其中$đ\xi = \dfrac{1}{(2\pi)^n}d\xi$，$p(x,\xi) = \sum\limits_{|\alpha|\leqslant m} a_\alpha(x)\xi^\alpha$，称为算子 P 的象征(或更准确地称为全象征). 它可以写成 $p_m(x,\xi) + p_{m-1}(x,\xi) + \cdots$，其中每一项都是 ξ 的某次齐次多项式. 然而就(11) 式中的积分来说，要求 $p(x,\xi)$ 对于 x 是相当一般的函数，而对于 ξ 只是多项式，这二者是不平等的. 因此，将 $p(x,\xi)$ 用关于 ξ 的更一般的函数代入，就提供了一种推广微分算子概念的可能性.

现在我们把象征类扩大，使其每一项不一定是 ξ 的多项式，而是类似于齐次函数的所谓 S^m 类函数，即满足条件

$$|\partial_\xi^\alpha\partial_x^\beta a(x,\xi)| \leqslant C(1 + |\xi|)^{m-|\alpha|}, \tag{12}$$

此外，和式中的项数也不一定有限，它可取为无限项这类函数的渐近和. 若有函数列$\{a_j(x,\xi)\}$，满足$a_j(x,\xi)\in S^{m_j}$，m_j下降趋于

$-\infty$，可取象征 $a(x,\xi)$ 为 $\{a_j(x,\xi)\}$ 的渐近和. 即对任意 k,

$$a(x,\xi)-\sum_{j=0}^{k-1}a_j(x,\xi)\in S^{m_k},\tag{13}$$

$\sum a_j(x,\xi)$ 即称为 $a(x,\xi)$ 的渐近展开，这里并不要求级数是收敛的.

记 $S^\infty=\bigcup\limits_m S^m$，将 S^∞ 作为象征类，引入

定义 2 若 $a(x,\xi)\in S^\infty$，则对 $\hat{u}\in C_0^\infty(\Omega)$，3 定义

$$Au=\int e^{ix\xi}a(x,\xi)\,\hat{u}\,(\xi)\,đ\xi,\tag{14}$$

称 A 为拟微分算子，并称 $a(x,\xi)$ 为 A 的象征.

当 $u(x)\in C_0^\infty(\Omega)$ 时，$\hat{u}\,(\xi)$ 在 $\xi\to\infty$ 时是速降的，所以(14)右端积分有意义，容易证明，$(Au)(x)$ 是 x 的 C^∞ 函数，而且 A 是 $C_0^\infty(\Omega)$ 到 $C^\infty(\Omega)$ 的连续映照.

为了运算的方便，我们常把(14) 改写为

$$Au=\iint e^{i(x-y,\xi)}a(x,\xi)u(y)\mathrm{d}y\,đ\xi$$

的形式，或更一般地考察积分

$$Bu=\iint e^{i(x-y,\xi)}b(x,y,\xi)u(y)\mathrm{d}y\,đ\xi,\tag{15}$$

其中 $b(x,y,\xi)$ 满足

$$|\partial_\xi^\alpha\partial_{xy}^\beta b(x,y,\xi)|\leqslant C(1+|\xi|)^{m-|\alpha|},\tag{16}$$

它也称为 S^m 类函数. 积分(16) 已经不是通常的 Lebesgue 积分，它可以按

$$\lim_{\varepsilon\to 0}\iint e^{i(x-y,\xi)}b(x,y,\xi)\alpha\left(\frac{|\xi|}{\varepsilon}\right)u(y)\mathrm{d}y\,đ\xi\tag{17}$$

来定义，其中 $\alpha(t)$ 是在 $|t|<1$ 时恒等于 1 的 C_0^∞ 函数. 当然，我们必须说明极限(17) 存在，且极限值与函数 α 的选取无关，而这些都是可以严格地证明的.

按(15) 式定义的 Bu 也称为拟微分算子，而且可以证明，在略去一个具 C^∞ 核的积分算子(以下称为光滑算子) 的差别下，它可

以表示为(14)的形式，这时，在形式(14)中出现的函数 $a(x,\xi)$ 仍然称为拟微分算子的象征.

拟微分算子有以下一些重要的性质.

一、拟局部性

当拟微分子算子 A 作用到一个函数 u 上时，不会扩大该函数的奇性集合，即

$$\text{sing supp } Au \subset \text{sing supp } u. \tag{18}$$

若在微局部意义下考察函数的奇性时，还有

$$\mathrm{WF}(Au) \subset \mathrm{WF}u. \tag{19}$$

二、拟微分算子构成代数

对于拟微分算子当然可以进行常规的加法与数乘运算. 此外，将算子复合作为乘法，拟微分算子集合构成一个代数，这里需要说明的是，为了使拟微分算子的定义域与值域一致，通常总是略去一个光滑算子的相差. 所以更严格地说，若以 $Ps(\infty)$ 记形为(15)的拟微分算子全体，$Ps(-\infty)$ 记光滑算子集，则 $Ps(\infty)/Ps(-\infty)$ 构成一个代数，而且这个代数中还有两种运算：转置与共轭.

有意思的是，对于拟微分算子的象征也可以规定某种运算，使象征全体按这种运算规则构成线性代数，并与拟微分算子代数构成一一对应. 与拟微分算子集合中视 $Ps(-\infty)$ 算子为零一样，在象征集合中也将 $S^{-\infty}=\bigcap_m S^m$ 中的任一元素视为零元素. 作为函数乘积求导法则中 Leibniz 公式的推广，我们定义"乘法"

$$q\overset{\circ}{\times}p \sim \sum_{\alpha}\frac{1}{\alpha!}\partial_\xi^\alpha q(x,\xi)D_x^\alpha p(x,\xi). \tag{20}$$

可以证明，以 $q(x,\xi)$ 为象征的算子 Q 与以 $p(x,\xi)$ 为象征的算子 P 的复合算子 $Q\circ p$ 恰具有象征 $q\overset{\circ}{\times}p$，又 $S^\infty/S^{-\infty}$ 关于加法、数乘以及"乘法"构成代数.

在象征集合中还可以引入“转置”与“共轭”两种运算，即

“转置”： $$\dot{t}(p(x,\xi)) \sim \sum_{\alpha} \frac{1}{\alpha!}\partial_{\xi}^{\alpha} D_{x}^{\alpha} p(x,\xi). \tag{21}$$

“共轭”： $$\dot{*}(p(x,\xi) \sim \sum_{\alpha} \frac{1}{\alpha!}\partial_{\xi}^{\alpha} D_{x}^{\alpha} p(\overline{x,\xi}). \tag{22}$$

也可以证明，以 $p(x,\xi)$ 为象征的拟微分算子 P 的转置 ${}^{t}P$ 以及共轭 P^{*} 恰具有如(21)(22)所示的象征，当然，这些都得在前述的不计 $Ps(-\infty)$ 类算子的差别以及不计 $S^{-\infty}$ 类象征的差别的意义上说的.

总之，$Ps(\infty)/Ps(-\infty)$ 与 $S^{\infty}/S^{-\infty}$ 成为同构，它们都是具有转置与共轭运算的线性代数.

三、微分流形上的拟微分算子

由于微分流形上分析学的需要，人们希望能够考察在微分流形上的拟微分算子，这一点是可以做到的. 首先，考察在欧氏空间坐标变换下所诱导出的拟微分算子及其象征的变化规律，设有 $\Omega_1 \to \Omega_2$ 的一个微分同胚 $\varphi: y = \varphi(x)$，则对于 Ω_2 上的函数 u，可得 Ω_1 上的函数 $\varphi^{*}u = u \circ \varphi$，今若给定拟微分算子 $A: C_0^{\infty}(\Omega_1) \to C^{\infty}(\Omega_1)$，则可以诱导出一个算子 $A_1 = \varphi_{*}A\varphi^{*}$，它是 $C_0^{\infty}(\Omega_2) \to C^{\infty}(\Omega_2)$ 的映照. 能够证明 A_1 也是拟微分算子，而且它的象征可以利用算子 A 的象征以及函数 φ 表示.

如果算子 A 的象征 $a(x,\xi)$ 有渐近展开 $\sum a_j(x,\xi)$，其中 $a_j(x,\xi) \in S^{m_j}$，$m_0 = \max\limits_{0 \leqslant j < \infty} m_j$，则称 $a_0(x,\xi)$ 为 $a(x,\xi)$ 的主象征. 由微分同胚 φ 所诱导出的拟微分算子 A_1 的主象征 $\sigma(y,\eta)$ 与 $a_0(x,\xi)$ 有关系

$$\sigma(y,\eta) = a_0(x,\xi)\big|_{x=\varphi^{-1}(y),\, \xi={}^{t}\varphi'(x)\eta}. \tag{23}$$

(23)式中所示的 (x,ξ) 与 (y,η) 之变换关系也正是余切丛的局部坐标的变换规律. 于是，以坐标变换下的拟微分子算子的变换规律为基础可以定义在微分流形上的拟微分算子，而此时拟微分算子

的主象征定义在该微分流形的余切丛上.

四、拟微分算子的有界性

当拟微分算子作用于 Sobolev 空间中的元素时，它常常具有某种有界性. 例如，若 A 为具 S^m 类象征的拟微分算子，则对 Ω 中任意紧集 K,K_1 有常数 C，使

$$\|Au\|_{H^s(K_1)} \leqslant C\|u\|_{H^{s+m}}, \forall\ u \in C_0^\infty(K). \tag{24}$$

(24) 式是指 H^s 空间中的有界性，通常称为 L^2 有界性，更进一步还可讨论 L^p 有界性，或在其他类型函数空间中的有界性或种种估计.

上述诸性质的证明以及关于拟微分算子其他性质的介绍可参考[1],[2].

利用拟微分算子为工具可以将有些定理的证明大大简化. 例如，椭圆型方程在右端是 C^∞ 函数的区域中解必为 C^∞ 的，这是一个著名的定理，现在利用拟微分算子可以很快地证明它. 设 P 是一个椭圆算子，函数 u 满足 $Pu = f$，在区域 ω 中 $f \in C^\infty(\omega)$，我们将设法找到 P 的一个"逆算子"Q，它满足 $QP - I = R$，其中 R 为光滑算子，如果这种拟微分算子 Q 存在，将 Q 作用于 $Pu = f$ 的两端，得

$$u + Ru = Qf. \tag{25}$$

由拟微分算子的拟局部性质知 $Qf \in C^\infty(\omega)$，而因 R 为光滑算子，故 $Ru \in C^\infty$，这样 u 在 ω 中就必定是 C^∞ 的，于是问题就归结为算子 Q 的寻求.

为寻求算子 Q 只需寻求它的象征，而若这个象征有一个渐近展开式，我们只需求出此渐近展开式的每一项，接着就可以用 Borel 技术构造出所需的象征，设 Q 的象征为 $\sum_0^\infty q_j$，算子 P 的象征为齐次多项式 $P_m(x,\xi)$（这并不影响方法的实质），比较 $QP = I +$

R 的象征的各同次幂,可得

$$
\begin{aligned}
& p_m(x,\xi)q_0(x,\xi) = 1, \\
& p_m(x,\xi)q_1(x,\xi) + \sum_{|\alpha|=1} \partial_\xi^\alpha q_0(x,\xi) D_x^\alpha p_m(x,\xi) = 0, \\
& \cdots\cdots \\
& p_m(x,\xi)q_k(x,\xi) + \sum_{\substack{|\alpha|+j=k \\ |\alpha|>0}} \partial_\xi^\alpha q_j(x,\xi) D_x^\alpha p_m(x,\xi) = 0, \\
& \cdots\cdots
\end{aligned}
\tag{26}
$$

由于 P 为椭圆算子,所以 $p_m(x,\xi) \neq 0$,从而从上式可逐项地决定 $q_0, q_1, \cdots, q_k, \cdots$,于是象征 $q(x,\xi)$ 以及算子 Q 均可求得,从而椭圆型方程解的内正则性获证.

上述关于椭圆算子的内正则性定理可以写成:若 P 为椭圆算子,则

$$
\text{sing supp } Pu = \text{sing supp } u. \tag{27}
$$

用类似的方法还可以证明

$$
\mathrm{WF}(Pu) = \mathrm{WF}u. \tag{28}
$$

这里的一个例子初步显示了拟微分算子的威力,在下节中我们将较多地叙述它的应用.

拟微分算子的理论到 70 年代又发展到了一个新的高度,Fourier 积分算子的系统理论建立了. Fourier 积分算子是拟微分算子概念的进一步推广,它将(15)式中指数函数 $e^{i\langle x-y,\xi\rangle}$ 替换成 $e^{i\varphi(x,y,\xi)}$,其中 $\varphi(x,y,\xi)$ 称为位相函数,满足一定的条件,例如关于 ξ 为齐一次,没有使 $\nabla\varphi = 0$ 的临界点等. Fourier 积分算子的一般形式是

$$
(Fu)(x) = \iint e^{i\varphi(x,y,\xi)} a(x,y,\xi) u(y) \mathrm{d}y \mathrm{d}\xi, \tag{29}
$$

其中记号 x, y, ξ 所表示的自变数的个数也不一定相等. 这类算子在讨论拟微分算子的化简,双曲型算子解的构造等问题中更为有效. 至于它的性质与应用的详述可参考有关著作,如[3].

回到本节之初所提出的微局部化问题，若在 $R_x^n \times R_\xi^n$ 中有一个单位分解 $1 = \sum \varphi_j(x,\xi)$，每个函数 $\varphi_j(x,\xi)$ 在 $\xi \to \infty$ 时是关于 ξ 零次齐次的（甚至是具有界支集的），则可以用 $\varphi_j(x,\xi)$ 为象征定义一个拟微分算子，记为 $\varphi_j(x,D)$. 于是，对于任意函数 $u(x)$，可以有分解

$$u(x) = \sum u_j(x) = \sum \varphi_j(x,D)u, \tag{30}$$

我们称每个单项 $u_j(x)$ 微局部地支于 supp φ_j 上. 这时，若有 S^∞ 函数 $a(x,\xi)$ 满足 supp $a \cap$ supp $\varphi_j = \varnothing$，则必有 $a(x,D)u_j \in C^\infty$.

§3　微局部分析的应用

用微局部分析的方法来研究偏微分方程导致了线性偏微分方程 的巨大进步. 人们借此对于偏微分方程理论的许多基本问题重新加以认识与处理，有些问题研究历史久长，而从此得到了迅猛的进展，也有些问题是近期才提出的，它们使偏微分方程理论研究达到了一个更深的层次. 我们几乎可以说，60 年代后线性偏微分方程 理论的每一个重要进展无一不是与微局部分析紧密相连的，以下举几个例子说明之.

一、柯西问题的唯一性

偏微分方程 Cauchy 问题唯一性的研究由来很久，早在 20 世纪初 Holmgren 证明了具有解析系数 Cauchy 问题在非解析函数类中解的唯一性. 但当方程系数仅为 C^∞ 函数时，Cauchy 问题不一定有唯一性. 到 30 年代末，对于含两个自变数方程的 Cauchy 问题唯性 才有了较普遍的结果，而对含多个变数的方程，直到 50 年代末以前仍只有零星的结果. 1958 年，Calderon 利用拟微分算子为工具证明了不具重特征的偏微分方程 Cauchy 问题解的唯一性，他只要求方程具 C^∞ 系数与不具有重特征，并不需要对方程的类型提

出什么要求.

Calderon 证明 Cauchy 问题唯一性的要点如下:他首先引入以 $|\xi|$ 为象征的拟微分算子(也称为奇异积分算子)将高阶方程组化成一阶方程组,这种化方程组的方法不会增加特征,因而所得到的方程组也不具有重特征,这个一阶方程组的象征是一个不具重特征值的矩阵,所以能够通过相似变换化成对角阵.然后问题大体上就相当于单个一阶方程的问题,从而较容易建立适当的估计式,而导致齐次 Cauchy 问题的解为零.这里需强调的是,我们将一阶方程组的象征矩阵化成对角阵的过程就相当于在拟微分算子范畴中对方程组进行变换,而且由于方程组的象征阵化成对角阵是在 $R^n_x \times R^n_\xi$ 的每点的邻域中进行的,因此每次所得到的估计都是微局部的.仅在将这些微局部的估计综合以后才得到一个整体的估计式.

二、高维双曲型方程组的初边值问题

在微局部分析理论出现以前,对高维双曲型方程组的初边值问题的研究通常采用能量积分的方法,这一方法对于对称双曲组比较有效,而对非对称双曲组就难以奏效.因为在高维的情形,一般不可能通过自变数与函数的变换把各个微分算子的系数阵同时化成对称阵,但是,从微局部的观点来看,整个微分算子组的象征只是**一个**矩阵,在每点 (x,ξ) 的邻域,它都可以通过矩阵的相似变换化成对称阵,翻译成拟微分算子的语言.就是说,原方程组可以微局部地化成对称的拟微分方程组.这样,能量积分法可以用到各个拟微分方程组上去,导得一系列的微局部的能量估计,综合之可以得到整体的能量估计,从而为获得初边值问题解的存在唯一性奠定了基础.此外,由于在选择上述的相似变换将原方程组的象征阵化成对称阵时有较多的灵活性,故有可能考虑取更一般的边界条件的初边值问题.60 年代,人们通常要求边界条件是"耗散"型

的，到 1970 年 Kreiss 对一般严格双曲组的研究中，仅要求边界条件满足 Lopatinski 条件.

三、局部可解性

大家知道，常微分方程总是有很多解的. 由于偏微分方程可以看 作是常微分方程的推广，人们长期认为一个偏微分方程也总是有很多解的，并需要以适当的定解条件来确定其解. 然而 1957 年 H. Lewy 给出了一个反例

$$\frac{\partial u}{\partial x} + i\frac{\partial u}{\partial y} - 2i(x+iy)\frac{\partial u}{\partial t} = f. \tag{31}$$

对这样的方程若仅要求 f 是原点邻域中的 C^∞ 函数，对于很多 f，方程根本就没有解. 这一新现象的发现，促使人们去研究这样一个问题，究竟什么样的方程才是对任意右端有解的. 更准确地说，若 L 为在 Ω 中定义的偏微分算子，$x_0 \in \Omega$，如果有 x_0 的邻域 ω，使对任一 $f \in C_0^\infty(\omega)$ 都有 $u \in \mathscr{D}'(\Omega)$，使 $Pu = f$ 在 ω 中成立，则称算子 L 在 x_0 点为局部可解的. 自 1957 年以后的一段相当长的时期中，局部可解性问题成了当时线性偏微分方程理论研究的一个热点.

1960 年 Hörmander 证明了实系数主型算子总是局部可解的，这里的主型算子也就是指没有重特征的算子. 1963 年 Nirenberg 与 Treves 对于具复系数的一阶主型算子局部可解的必要条件与充分条件得到了较完整的结果. 但是对于高阶主型算子的局部可解性直到 1970 年才有较完整的结果，其原因也是在此以前缺乏合适的分析工具. 1970 年 Nirenberg 与 Treves 的做法是先在高阶算子的特征点邻域中将它分解成一个椭圆算子与一个一阶算子的乘积，后者在该邻域中与原高阶算子具有相同的特征点，然后集中研究 该一阶算子的局部可解性. 由于这种算子分解只能在拟微分算子 的范畴中进行，因而也只有在拟微分算子的系统理论发展与成

熟以后，局部可解性问题的研究才有了重大的突破．然而，应该指出，这个问题至今仍未彻底解决．

四、奇性传播

偏微分方程解的正则性是偏微分方程理论研究的基本课题之一，正则性与奇性实际上是同一对象的两个侧面，因此人们一般更注重于偏微分方程解的奇性生成、分布、强度或奇性结构．关于偏微分方程解的奇性分布的一个经典的结论是：解的弱间断只能在特征曲面上出现，较精细的结论是，解的弱间断是沿方程的次特征线传播的，对于线性主型偏微分方程，则有更精确的奇性传播定理，它是 Hörmander 等人在 60 年代末建立的．

设 P 为具 C^∞ 系数的主型算子，它的主象征为 $p(x,\xi)$，称 Hamilton 方程组

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}s} &= \frac{\partial p}{\partial \xi}, \\ \frac{\mathrm{d}\xi}{\mathrm{d}s} &= -\frac{\partial p}{\partial x} \end{aligned} \tag{32}$$

满足 $p(x(s),\xi(s))=0$ 的解 $x(s),\xi(s)$ 为算子 P 的零次特征带，则有

定理 若 $u\in\mathscr{D}'(\Omega)$ 为方程 $Pu=f$ 的解，其中 $P=p(x,D)$ 如上述，$p(x_0,\xi_0)=0$，$(x_0,\xi_0)\in \mathrm{WF}u$，$\gamma$ 为过 (x_0,ξ_0) 的零次特征带，则若 $\gamma\notin\mathrm{WF}f$，必有 $\gamma\notin\mathrm{WF}u$．

上述定理也可表述为，在微局部的意义下，当 f 为光滑时，只要 u 在某点有奇性，这个奇性就沿着次特征带传播，显然，这一结论比经典的在底空间中的奇性传播结论更精细，而通过往底空间的投影，很容易导出经典的奇性传播定理．

奇性传播的研究完全是属于微局部分析范畴的课题，它引导人们对偏微分方程解的性质作更深入的了解．

关于微局部分析对偏微分方程研究的影响还可举出很多例子，例如Cauchy问题适定性的研究，非椭圆方程解的次椭圆估计，重特征算子的分类与化简，等等.此外，它在散射理论、量子力学的准经典近似等方面的应用也是令人瞩目的.

五、非线性微局部分析

微局部分析在线性偏微分方程理论研究中所取得的出色成就必然吸引人们将微局部分析方法应用于其他各种问题，特别是非线性偏微分方程的研究.近十年来在这一方面人们也取得了很大的成功，从而形成了非线性微局部分析的分支.其中最重要的成果有非线性方程解的奇性传播以及高维守恒律双曲型方程组广义解的存在性等.下面我们分别简述之.

由于非线性效应，方程的解的奇性会产生相互干扰，因而一般来说，它的解的奇性分布比相应的线性方程解的奇性分布要复杂得多.目前，人们通过多年的研究对非线性方程解的奇性传播、反射、干扰等规律已有了较清楚的认识.特别是法国J.M.Bony在研究这类问题中又发展了仿微分算子的理论与二次微局部分析的理论，从而为微局部分析的应用开辟了更广阔的前景.

在高维守恒律双曲型方程组的研究中，带有各种类型间断的解的存在性等问题长期以来是人们十分向往而又束手无策的问题.因此，尽管一维守恒律双曲型方程组的研究已有长久的历史与很好的精细结果，但一到高维的情形，很多结论是否成立即成为未知.现在鉴于微局部分析理论的发展，人们能够先对相应的线性化问题作出比以往更精细的估计，从而导得非线性问题的所需结果.例如，在80年代，A.Majda与S.Alinhac先后获得了高维守恒律双曲型方程组带有激波与带有中心波的解的存在性，相应地还有一批结果涌现，从而使人们在这个领域中的研究获得了重大的突破.当然，在高维情况下，非线性守恒律双曲组广义解研究的内容十分

丰富,还有许多复杂而深刻的问题未解决.

近年来,J. Y. Chemin 利用微局部分析对不可压缩流体的数学理论进行研究,也取得了重要的进展.

目前,作为一个新兴的学科分支,微局部分析还在迅速地发展,它的应用也正在渗透到更多的方面.它无疑是当代数学发展的一个主攻方向,更丰硕的成果尚有待我们努力去摘取.

参 考 文 献

1. 齐民友,线性偏微分算子引论,科学出版社,1986.
2. 陈恕行,偏微分方程概论,高等教育出版社,1981.
3. 仇庆久,陈恕行,是嘉鸿,刘景麟,蒋鲁敏,傅里叶积分算子理论及其应用,科学出版社,1985年.
4. 陈恕行,仇庆久,李成章,仿微分算子引论,科学出版社,1990.
5. L. Nirenberg,线性偏微分方程讲义,陆柱家译,上海科学技术出版社,1980.
6. L. Hörmander, The analysis of linear partial differential operators, I — IV, Springer-Verlag, 1985.
7. L. Nirenberg and F. Trèves, On local solvability of linear partial differential equations. Part I: Necessary conditions. Part II: Sufficiently conditions. Comm. Pure Appl. Math. v. 23, 1970.
8. J. M. Bony, Calcul symbolique et propagation des singularites pour les equations aux derivees partiel les non-lineaires. Ann. Scient, F. N. S. ,v. 14,1981.
9. J. M. Bony, Second microlocalization and propagation of singularities for semilinear hyperbolic equations, "Contribution"Taniguchi Symp. HERT,Katata 1984".
10. A . Majda , The stability of multi - dimensional shock fronts , The existence of multidimensional shook fronts, Mem. Amer. Math. Soc. v. 275 & v. 283,1983.

11. S . Alinhac , Existence d'ondes de rarefaction pour systemes quasi-lineaires hyperboliques multidimensionnels, Comm. in PDEs, v. 14 (2),1989.

12. J. Y. Chemin, Une facette mathématique de la mécanique des fluides I , Publications of Ecole Polytechnique No. 1055, 1993.

Ⅲ

动力系统
Dynamical Systems

张景中
(中国科学院成都计算机应用研究所)
Zhang Jingzhong
(Chengdu Institute of Computer Applications, Academia Sinica)

微分动力系统——时间的数学
Differential Dynamical Systems ——The Mathematics of Time

§1　决定性系统的数学模型

我们把不断变化发展着的一件事物叫做一个系统.系统可以很简单,例如一个自由下落的质点;也可以很复杂,例如银河系、大型计算机.在一定时刻 t,系统的状态用 X_t 表示.随系统的不同,X_t 也许是一个实数,也许是一组实数,也许是曲面上的一个点,也许是一个无穷数列,也许是一条曲线或一张曲面.一系统所有可能状态的集合记作 M.

如果一个系统的变化发展完全由内部的因果关系所制约,就叫做决定性系统.对这样的系统,只要知道了它在某一时刻 t_0 的状态,就能准确地预测它未来的命运并回溯它的历史发展过程.当然,这要假定我们有足够的知识.

用数学语言来描述,决定性系统是这样的系统:在取定了计时起点之后,系统在时刻 t 的状态 X_t 完全由初始状态 X_0 与时间 t 所确定.用符号表示就是说:$X_t = F(t,X_0)$,即 X_t 是 t 和 X_0 的函数.而且这个函数 F 与计时起点的选取无关.

但并非随便什么函数都可以充当这种描述决定性系统因果关系的工具.容易看出,F 应当满足下列两个条件:

$$\begin{cases} 1^\circ \quad F(0,X) = X; \\ 2^\circ \quad F(t,F(s,X)) = F(s+t,X) \\ \qquad (X \in M;\ s,t \in (-\infty + \infty)). \end{cases} \tag{1}$$

这两条的意思直观上很清楚:1° 表示,时刻为 0 的状态是初始状态;2° 表示,若从状态 X 出发,经过时间 s 达到了状态 $Y = F(s, X)$;从状态 Y 出发,经过时间 t 达到状态 Z;则从 X 出发,经过时间 $s+t$ 也达到状态 Z.这是说:因果关系与时间的起点无关,只决定于时间间隔.

一般来说,给了状态集 M 和满足(1)的一个映射

$$F: \quad R \times M \to M \tag{2}$$
$$(Y = F(t,X), t \in R; X, Y \in M)$$

（这里 R 即全体实数集$(-\infty, +\infty)$），我们便说有了一个 M 上的动力系统 $F(t,X)$.

至于动力系统这个名目，是因为当初 Poincaré 采用了这样的数学模型研究多个天体（例如日、月、地球三体）在万有引力作用下的运行规律. 这正好是一个质点组动力学系统.

在(2)中，M 可以是区间，或平面、空间区域、曲线、曲面或多维的流形. 通常要求 F 对 t 和 X 可微，这时称系统 $F(t,X)$ 为微分动力系统. 如果变量 t 只能取非负实数，$t \in (o, +\infty)$，这意味着从初始状态出发我们只能预测未来而无法回溯历史，这时叫半动力系统.

我们可以从银幕上选取一系列画面编成电影故事连环画. 这个办法也可以用来研究动力系统. 在(2)中取 $t = 1$，便得到一个 M 到自身的映射 —— 短期因果关系映射：

$$f: \quad M \to M \tag{3}$$
$$(Y = f(X) = F(1,X)).$$

由(1)可知

$$\left.\begin{aligned} &f(f(X)) = F(1,F(1,X)) = F(2,X), \\ &\underbrace{f(f(\cdots f}_{n\text{重}}(X)\cdots)) = F(n,X). \end{aligned}\right\} \tag{4}$$

这表明：映射 f 的迭代产生的序列，可以用来描述系统在时刻 $t = 1,2,3,\cdots\cdots$ 处的一串画面. 因此，我们把 M 上的自映射 f 所产生的迭代序列：

$$\begin{cases} f^0(X) = X, \\ f^{n+1}(X) = f(f^n(X)) \end{cases}$$
$$(n = 0,1,2,\cdots) \tag{5}$$

叫做 M 上的离散半动力系统. 如果 F 可逆,用 f^{-n} 记其逆映射 f^{-1} 的 n 次迭代,则在(5) 中 n 可以取全体整数,“半”字就可以去掉了.

由于近几十年来电子计算机技术的迅速发展,使迭代的实际计算变得轻而易举,因而离散动力系统的研究十分活跃. 于是人们常常把离散动力系统就叫做动力系统. 同时把由(1) 所定义的动力系统 $F(t,X)$ 叫做 M 上的一个流.

除了流和迭代之外,描述决定性系统的又一重要数学模型是微分方程.

对流 $F(t,X)$,令

$$X(t) = F(t,X_0), \tag{6}$$

并将 $X(t)$ 对 t 求导:

$$\begin{aligned}\dot{X}(t) &= \frac{\mathrm{d}X(t)}{\mathrm{d}t} = \lim_{h\to 0}\frac{X(t+h)-X(t)}{h}\\ &= \lim_{h\to 0}\frac{F(t+h,X_0)-F(t,X_0)}{h}\\ &= \lim_{h\to 0}\frac{F(h,F(t,X_0))-F(t,X_0)}{h}\\ &= \lim_{h\to 0}\frac{F(h,X(t))-X(t)}{h},\end{aligned} \tag{7}$$

最后的极限仅与 $X(t)$ 有关,因而它可以表成状态 $X(t)$ 的函数 $P(X)$. 于是得到:

$$\dot{X} = p(X). \tag{8}$$

这是一个右端不包含主变元 t 的常微分方程组. 通常称之为定常的常微系统. 它从另一角度描述了决定性系统.

因此,微分动力系统与常微分方程便结下了不解之缘. 不过,本文的目的只是介绍迭代生成的动力系统的一点有关的基本知识. 更进一步的了解可看[1],[8],[9].

§2 一维迭代的若干简单实例

微分动力系统的研究，已运用着相当深奥的数学工具，产生了大批艰深的结果. 但也有些极其有趣的发现，却是不用多少数学知识的预备便可品尝的. 这主要限于一维情形.

通常，一个很简单的函数 f，它的 n 次迭代 f^n 可能变得很复杂. 例如，$f(x) = 1 - x^2$ 时，f^{10} 是 1024 次多项式. 而 f^{20} 的表达式可以印成厚厚的一本大书！

但也有些函数，迭代起来相当简单. 如：

例 1 $f(x) = x + a, f^n(x) = x + na$

(对应地有流 $F(t,x) = x + at$).

例 2 $f(x) = \lambda x, f^n(x) = \lambda^n x$

(对应地有流 $F(t,x) = \lambda^t x$).

例 3 $f(x) = \lambda x + a, f^n(x) = \lambda^n x + \dfrac{1-\lambda^n}{1-\lambda}a$

$\left(\text{对应地有流 } F(t,x) = \lambda^t x + \dfrac{1-\lambda^t}{1-\lambda}a\right)$.

例 4 $f(x) = \dfrac{x}{1+bx}, f^n(x) = \dfrac{x}{1+nbx}$

$\left(\text{对应地有流 } F(t,x) = \dfrac{x}{1+tbx}\right)$.

例 5 $f(x) = x^2, f^n(x) = x^{2^n}$

(对应地有流 $F(t,x) = x^{2^t}$).

从这些简单的迭代出发，使用相似变换的技巧，可以写出更多的迭代表达式[2]. 设想我们已知道了 f^n 的表达式，但不知道 g^n 的表达式. 如果能找到一个可逆的映射 $h(x)$，使

$$g = h^{-1} \circ f \circ h$$

$$(\text{即 } g(x) = h^{-1}[f(h(x))]), \tag{9}$$

则立刻得到：

$$g^n(x)=h^{-1}[f^n(h(x))], \tag{10}$$

这就找到了 g^n 的表达式.

例 6 $g(x)=2x^2-1$，取 $f(x)=2x$，

令 $h(x)=\arccos x$，得

$h^{-1}[f(h(x))]=\cos 2(\arccos x)$

$=2\cos^2(\arccos x)-1=2x^2-1=g(x)$，

故 $g^n(x)=\cos[2^n(\arccos x)]$.

例 7 $g(x)=\dfrac{2x}{1-x^2}$，取 $f(x)=2x$，

令 $h(x)=\operatorname{arc\,tg} x$，得

$h^{-1}[f(h(x))]=\operatorname{tg}2(\operatorname{arc\,tg} x)$

$=\dfrac{2\operatorname{tg}(\operatorname{arc\,tg} x)}{1-\operatorname{tg}^2(\operatorname{arc\,tg} x)}=\dfrac{2x}{1-x^2}=g(x)$，

故 $g^n(x)=\operatorname{tg}(2^n\operatorname{arc\,tg} x)$.

例 8 $g(x)=\dfrac{x}{\sqrt[k]{x^k+1}}$，取 $f(x)=\dfrac{x}{x+1}$，

令 $h(x)=x^k$，得

$h^{-1}[f(h(x))]=\dfrac{x}{\sqrt[k]{x^k+1}}=g(x)$，

故 $g^n(x)=\dfrac{x}{\sqrt[k]{1+nx^k}}$.

例 9 $g(x)=\dfrac{x^2}{2x-1}$，取 $f(x)=x^2$，

令 $h(x)=\dfrac{x}{x-1}$，得

$h^{-1}(f(h(x))=\dfrac{x^2}{x^2-(x-1)^2}=\dfrac{x^2}{2x-1}=g(x)$，

故 $g^n(x)=h^{-1}[(h(x))^{2^n}]=\dfrac{x^{2^n}}{x^{2^n}-(x-1)^{2^n}}$.

但是，这种方法的成功大体上有赖于巧合. 事实上，能写出 n

次迭代的具体表达式的函数实在是凤毛麟角．对于绝大多数写不出表达式的迭代映射产生的动力系统，就要通过其他方式来研究序列

$$X_0, X_1, X_2, \cdots, X_n, \cdots \quad (X_{n+1} = f(X_n)) \tag{11}$$

的变化趋势．这样的序列，叫做在 f 作用下过 X_0 的一条正向轨道．

如果 f 是区间上的自映射，就有一种浅显直观的几何图示法来描述轨道．请看图 1：在直角坐标系中画出函数 $f(x)$ 的图象 Σ 和直线 $l: y = x$，在 x 轴上任取一点 x_0，自 x_0 引 Y 轴的平行线交 Σ 于 P_0，则 P_0 的纵坐标为 $x_1 = f(x_0)$．自 P_0 引 X 轴的平行线交 l 于 Q_0，Q_0 的横坐标即 x_1．自 Q_0 又引 Y 轴平行线交 Σ 于 P_1，则 P_1 的纵坐标恰好是 $x_2 = f(x_1)$，…，这样下去一步一步画出一张蛛网图，从图上可以看出 $x_0, x_1, x_2 \cdots$ 这条轨道的来龙去脉．

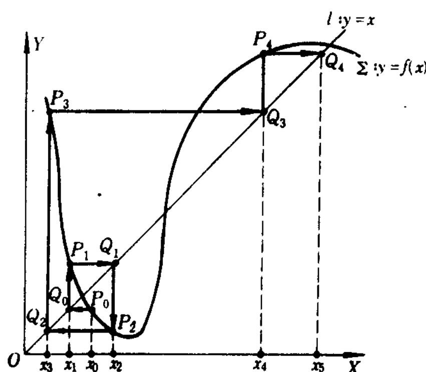


图 1

这种蛛网图，确实能对一些实际问题提供有价值的咨询．下面是两个应用实例：

例 10 设鱼塘里的鱼一年长成，并且在繁殖若干代之后自然死亡．于是在一定外界条件下，今年塘中成鱼的数量可以决定明年成鱼的数量，亦即明年成鱼数量 y 与今年成鱼数量 x 之间有着某种函数关系 $y = f(x)$．经过一番调查研究，这种函数的图象大体上如图 2．

从图 2 看出，曲线 Σ 与直线 l 有一个交点 $P=(x^*,x^*)$. 这个 x^* 满足 $f(x^*)=x^*$. 凡满足方程

$$f(x)=x \tag{12}$$

的 x 叫做 f 的不动点. 这里不动点 x^* 意味着鱼群数量的平衡点：如果今年有这么多鱼，明年也这么多 —— 水面负荷已达到了极限. 如果初始鱼群 $x_0\in(0,x^*)$，则由图可见 $x_1,x_2,x_3,\cdots$ 将一年年增长而趋于平衡点. 而当 $x_0'>x^*$ 时，$x_1',x_2',\cdots$，将一年年减少而趋于平衡点.

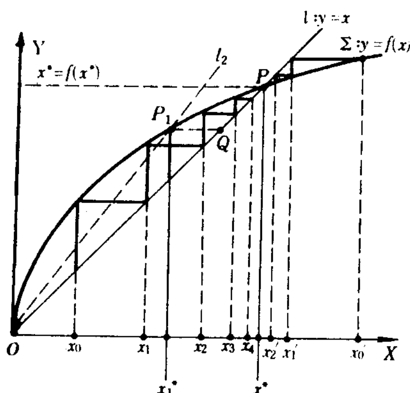


图 2

图上还有另一个不动点 $x=0$，很显然，这表示着今年塘里无鱼，明年也不会有. 但 $x=0$ 与 $x=x^*$ 这两个不动点有明显的不同. 当初始值稍稍偏离 $x=0$ 时，以后会一年年离 $x=0$ 更远. 反之，初始值稍稍偏离 $x=x^*$ 时，则会一年一年地更趋近回来. 所以，把 $x=0$ 这样的不动点叫做不稳定不动点，而 $x=x^*$ 叫做稳定不动点.

如果 $\bar{x}$ 是 f 的不动点，则当 $|f'(\bar{x})|>1$ 时它是不稳定的，$|f'(\bar{x})|<1$ 时是稳定的. $|f'(\bar{x})|=1$ 时两种情形都会发生.

图 2 中有一条斜的虚线 l_2，叫捕捞参考线. l_2 的斜率可由捕捞工作量确定. l_2 与 Σ 的交点 P_1 的横坐标 x_1^* 是由于捕捞而形成的新的平衡点. 捕捞量愈大，x_1^* 愈小. 竭泽而渔，则 x_1^* 趋于 0. 横线 P_1Q 反映渔获量的大小. 从图上看出，捕捞过多过少都是不利的.

这里把复杂事情简单化了. 因为决定下一年鱼群数量的因素，除今年鱼群数量之外，还有群中鱼的年龄分布. 那样就不能用一元

函数描述了.

类似的方法，可用来研究人口增长的规律，生物种群的消长等.

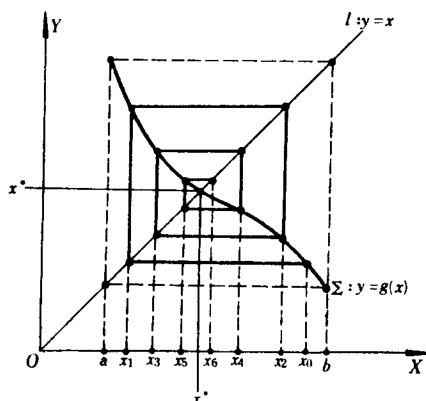


图 3

例 11 很多种果树的收获量有大小年之分. 这种现象也可以用迭代的蛛网图来描述.

在一定条件下，我们可以设想果树明年的产量 y 是今年产量 x 的函数：$y=g(x)$. 但它与鱼塘问题的函数有不同的特征.

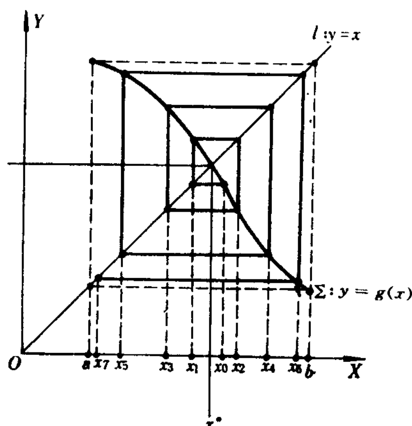


图 4

图 3 表示某种大小年现象不显著的果树. 从大年 x_0 出发，接着是小年 x_1，大年 x_2，小年 x_3，…. 但大小年的差别越来越小，最后稳定于不动点 x^* —— 常年产量. 但图 4 则是另一种状况：即使前两年 x_0 与 x_1 相差甚微，但年复一年地差距愈来愈大，最后趋于两点 a、b. 这里 $g(a)=b, g(b)=a$. 两点放在一起叫做 g 的一个"2-周期轨". 而 $x=a$ 和 $x=b$ 都叫做 g 的二周期点. 具有这种曲线的果树，大小年现象将是严重的，应采取适当园艺措施.

在图 3 与图 4 中，迭代序列

$$x_0, x_1, x_2, \cdots, x_n, \cdots \tag{13}$$
$$(x_{n+1} = g(x_n))$$

随 n 的增长而上下起伏振动，这叫做"非线性振荡". 非线性振荡现象在电磁学、力学、物理、化学中到处出现. 自然界的昼夜交替、寒来暑往、生物繁殖、波浪起伏等司空见惯的周期变化也可归结为非线性振荡. 人们对非线性振荡的研究已上百年. 微分动力系统这一研究领域的兴起，为非线性振荡研究提供了新的思路与方法.

§3 Sarkovskii 序与 Feigenbaum 现象

前面在鱼塘问题中遇到了不动点 $f(x^*) = x^*$，在果树问题中又遇到了 2- 周期点 $a = g(b)$ 和 $b = g(a)$. 推而广之，可以引入更一般的周期点概念：如果 x_0 满足

$$\begin{cases} 1^\circ \quad f^n(x_0) = x_0, \\ 2^\circ \quad \text{对任意 } 0 < k < n, f^k(x_0) \neq x_0 \end{cases} \tag{14}$$
$$(n、k \text{ 是正整数}),$$

则称 x_0 是映射 f 的一个 n- 周期点. 这时，称点组 $\{f^k(x_0); k = 1, 2\cdots, n\}$ 为 f 的一个 n- 周期轨.

周期轨可以表示自然现象或技术过程中出现的周期状态，在微分动力系统的研究中占有重要地位. 在许多有关周期轨的研究成果中，容易理解而又极其有趣的应当说是 60 年代发现的 Sarkovskii 定理了.

利用连续函数的介值定理容易证明：若 f 是闭区间$[a,b]$上的连续函数而且对一切$[a,b]$中的 x 有 $a \leqslant f(x) \leqslant b$，则必有 $x^* \in [a,b]$ 使得 $f(x^*) = x^*$. 这肯定了 1- 周期轨的存在条件.

对 2- 周期轨，有人证明过一个有趣的结果：设 f 是闭区间$[a,b]$上的连续自映射，则 f 没有 2- 周期轨的充分必要条件是：对一

切 $x \in [a,b]$，$\lim\limits_{n\to+\infty} f^n(x)$ 存在. 这表明：没有 2- 周期轨的连续函数，也不会有 $n(>2)$ 的周期轨. 它只能是相当“简单”的函数.

那么，有各种周期轨的函数是不是一定很复杂呢？倒不一定. 图 5 是一个很简单的函数的图象. 一眼看出它有一个 3- 周期轨：

$$\left\{f(0)=\frac{1}{2}, f\left(\frac{1}{2}\right)=1, f(1)=0\right\}.$$

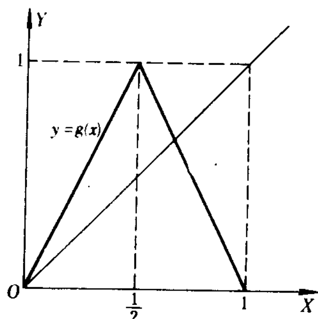


图 5

它有没有别的周期轨呢？有没有 4- 周期轨？11- 周期轨？108- 周期轨？似乎很难回答. 但是，只用很少数学知识（连续函数的介值定理就够了）便能证明下列不平常的事实.

“若在区间上定义的连续函数 f 有 3- 周期轨，则对任一正整数 n，f 有 n- 周期轨.”

比这个结果更有趣和更细致的，是著名的 Sarkovskii 定理. 为叙述这一定理，先把全体自然数重新排个顺序. 用记号 $m \lhd n$ 表示“m 先于 n”，则 Sarkovskii 序为：

$$3 \lhd 5 \lhd 7 \lhd 9 \lhd 11 \lhd \cdots \lhd 2n+1 \lhd \cdots$$
$$6 \lhd 10 \lhd 14 \lhd 18 \lhd \cdots \lhd 2(2n+1) \lhd \cdots$$
$$2^2 \cdot 3 \lhd 2^2 \cdot 5 \lhd 2^2 \cdot 7 \lhd \cdots \lhd 2^2(2n+1) \lhd \cdots$$
$$\cdots\cdots$$
$$2^k \cdot 3 \lhd 2^k \cdot 5 \lhd 2^k \cdot 7 \lhd \cdots \lhd 2^k(2n+1) \lhd \cdots$$
$$\cdots\cdots \lhd 2^{l+1} \lhd 2^l \lhd \cdots \lhd 16 \lhd 8 \lhd 4 \lhd 2 \lhd 1.$$

也就是说，全体大于 1 的奇数自小而大排在最前面，接着是它们的 2 倍，4 倍，8 倍，…，2^k 倍，…，最后是所有 2 的方幂，但顺序是由大到小，最后是 1. 按此顺序，有：

Sarkovskii 定理 设 f 是区间上的连续函数. 如果 f 有 n- 周

期点，则当 $n \lhd m$ 时 f 一定有 m- 周期点.

这条定理发表于 60 年代. 近年来它引起了广泛的兴趣和一系列有关的研究. 其中特别值得一提的是所谓 Feigenbaum 现象.

设在 $[-1,1]$ 上定义了一个单参数的单峰函数族 $f_\mu(x)$，参数 $\mu \in [0,a]$，使得(如图 6)：

1° $f_\mu(x)$ 对变元 x 和参数 μ 连续可微.

2° $f_\mu(0) = 1$, $f_\mu(-1) = f_\mu(1) \leqslant 1$.

3° $f_\mu(-x) = f_\mu(x)$.

4° 对固定的 μ，f_μ 在 $[0,1]$ 上单调下降.

5° $f_a(-1) = f_a(1) = -1$.

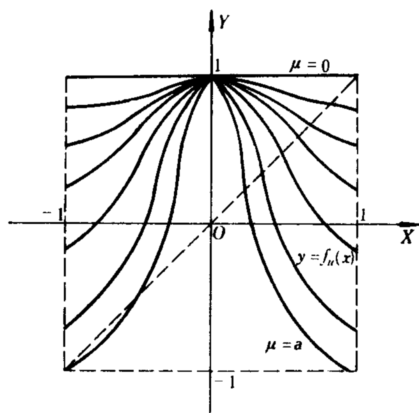


图 6

则当参数 μ 由 0 逐渐增大时，f_μ 将具有一切 n- 周期点. 设 μ_n 是使 f_{μ_n} 具有 n- 周期点的最小参数(即当 $\mu < \mu_n$ 时 f_μ 即无 n- 周期点)，则当 $m \lhd n$ 时必有 $\mu_n < \mu_m$. 这一结果把 Sarkovskii 序和单参数函数族联系起来了.

Feigenbaum 研究了与昆虫数量增减有关的所谓虫口差分方程

$$x_{n+1} = 1 - \mu x_n^2 \quad (0 \leqslant \mu \leqslant 2, x_n \in [-1,1]), \qquad (15)$$

这实际上是函数族

$$f_\mu(x) = 1 - \mu x^2 \quad (\mu \in [0,2], x \in [-1,1]) \qquad (16)$$

的迭代问题. 如果对固定的 μ，f_μ 具有超稳定 n- 周期点，就表明对应于参数 μ 的昆虫数量变化以 n 年为周期. 所谓超稳定 n- 周期点，是指 f 的周期点 x_0，它同时满足

$$(f^n(x))'_{|x_0} = 0\,. \tag{17}$$

超稳定周期点对应于易于观察到的周期现象，在计算机上也极易计算，故特别受到重视.

Feigenbaum具体地计算出使f_μ有超稳定的2^n-周期点的一系列μ值：$\mu_1,\mu_2,\mu_4,\mu_8,\cdots,\mu_{2n},\cdots$. 他发现当$n$越来越大时有：

$$\frac{\mu_{2^n}-\mu_{2^{n-1}}}{\mu_{2^{n+1}}-\mu_{2^n}} \longrightarrow \delta = 4\cdot 669201609\cdots. \tag{18}$$

当用另一个单峰函数族代替$f_\mu(x)=1-\mu x^2$时，居然还得到这同一个常数δ. 不仅如此，这个常数δ竟又一再在许多物理问题中出现. 它是一个“普适常数”！这一现象引起了广泛的兴趣. 相继而起的研究形成了一个十分活跃的领域.

在数学和自然科学中，有一个最重要的常数π，它几乎到处出现. 再一个重要常数是e，它在微积分学中扮演了重要的角色. 人们说，在Feigenbaum现象中出现的这个δ，也许可以叫做第三个常数了吧！（也有人认为第三常数是黄金分割数$\frac{1}{2}(\sqrt{5}-1)=0.618\cdots$；而$\delta$为第四常数.）

对Feigenbaum现象的研究已成为一个活跃的方向. 这一现象的数学机理尚未得到彻底的揭示.

与本节有关的更详细的内容，参看[4]，[5].

§4 结构稳定与双曲性

在前面已经提到，动力系统研究的基本课题，是根据短期的因果关系映射$Y=F(X)$的性质去了解系统的长期变化状态，即轨道：

$$X_0,X_1,X_2,\cdots,X_n,\cdots \tag{19}$$
$$(X_{n+1}=F(X_n))$$

的性质. 这就很自然地产生了一个基本问题：根据短期因果关系，

能否有效地预见系统的长远命运?

这个问题是有道理的.如果我们研究的是实际出现的自然现象或科学技术过程,由于测量手段的限制或理论的简化,所写出的短期因果关系映射 $Y = F(X)$ 只能是实际过程的近似描述.尽管它可以十分精确地与真实情形近似,但毕竟不能绝对一致.那么,在无限迭代的过程中,误差会不会越来越大,以至"差之毫厘,谬以千里",最终得出的结论与实际过程竟大不相同呢?

确实有可能这样,请看图 7 中的三个图象:

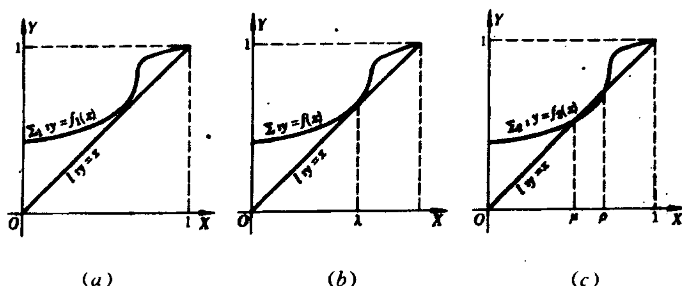


图 7

在图 7(b) 中画出一个[0,1]上的连续可微函数 $f(x)$ 的图象,它的迭代序列性质是:

1_b:有两个不动点 $x_1 = \lambda, x_2 = 1$;

2_b:若 $x \in (0,\lambda)$,则序列$\{f^n(x)\}$ 递增地趋于 λ;

3_b:若 $x \in (\lambda,1)$,则序列$\{f^n(x)\}$ 递增地趋于 1.

把 $f(x)$ 稍微扰动一下,变成相差极小的 $f_1(x)$,如图 7(a) 所示,其迭代序列会有本质的不同:

1_a:f_1 有一个不动点 $x = 1$;

2_a:对一切 $x \in [0,1)$,序列$\{f_1^n(x)\}$ 递增地趋于 1.

把 $f(x)$ 按另一种方式扰动一下,变成图 7(c) 中的 $f_2(x)$,迭代序列的性质又变了:

1_c：f_2 有三个不动点：$x_1 = \mu, x_2 = \rho, x_3 = 1$.

2_c：当 $x \in (0,\mu)$，序列 $\{f_2^n(x)\}$ 递减地趋于不动点 $x = \mu$.

3_c：当 $x \in (\mu,\rho)$，序列 $\{f_2^n(x)\}$ 递减地趋于不动点 $x = \rho$.

4_c：当 $x \in (\rho,1)$，序列 $\{f_2^n(x)\}$ 递减地趋于不动点 $x = 1$.

例子表明，多么微小的误差也会招致迭代序列的明显的性质改变.

对什么样的映射 F，小的扰动不会改变它的迭代序列的基本特性呢？这引出了结构稳定的概念：

定义 设 M 是 Riemann 流形（可理解为上面定义了距离的曲线或曲面），映射

$$F: M \to M \tag{20}$$

是连续可微的. 如果存在 $\varepsilon > 0$，使得对任一个 F 的小扰动 F^*，只要

$$\max_{X \in M}\{\rho(F(X),F^*(X)) + \|DF(X) - DF^*(X)\|\} < \varepsilon, \tag{21}$$

便存在 M 到自身的同胚 h（即 M 上可逆的连续映射），满足

$$1^\circ \quad F^*(X) = h^{-1}[F(h(X))]; \tag{22}$$

$$2^\circ \quad \max_{X \in M}\{\rho(X,h(X))\} < \varepsilon, \tag{23}$$

便说 F 是结构稳定的.

这里，$\rho(X,Y)$ 表示 M 上两点 X、Y 之间的距离，而 $DF(X)$ 是 F 在 X 处的微分 —— 可理解为映射 F 在 X 处的 Jacobi 阵 —— 偏导数矩阵.

条件 $F^* = h^{-1} \circ F \circ h$ 我们在第 2 节(9)式中已见过. 当有连续可逆的 h 使这等式成立时，便说 F^* 与 F“拓扑共轭”. 拓扑共轭是一种等价关系. 互相拓扑共轭的映射，它们有相同的动力系统性质：例如，有相同个数的周期轨. 这是因为 $F^{*n} = h^{-1} \circ F^n \circ h$ 之故.

因此，简单地说：如果映射 F 在足够小的 C^1 扰动下仍和未扰动前拓扑共轭，则称 F 是结构稳定的.

所谓C^r扰动，是指扰动之后得到的F^*不仅使$F^*(X)$与$F(X)$相差很小，连偏导数DF^*和DF也相差很小.

对于结构稳定的映射F，足够小的扰动不会带来其迭代序列性质原则上的变化，因此，我们可从它出发研究其迭代序列的性状而不担心失去对实际系统的有效性.

因此，弄清什么样的映射(首先是可逆的映射 —— 同胚)才具有结构稳定性，便成了微分动力系统的中心问题之一.

经过一些数学家们出色的工作，基本上已弄清楚：结构稳定性的基本条件是只具有双曲周期点，即所谓双曲性.[1]

回顾图7的(b)，$f(x)$扰动后发生不稳定现象，问题主要出在不动点$x=\lambda$处. 这个点是曲线Σ与直线$y=x$的切点，$f'(\lambda)=1$. 这启发我们：在不动点处导数为1，可能是不稳定的根源！

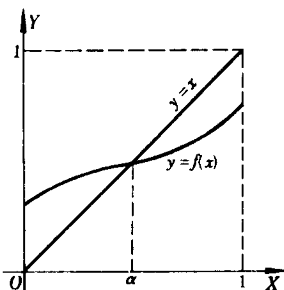


图 8

这个想法基本上猜对了. 像图8那样的可逆函数$f(x)$，它只有一个不动点α，而且在这点的导数$0<f'(\alpha)<1$，它确是结构稳定的.

对于区间上的函数$f(x)$，一般而言，如果在其n-周期点x_0处有

$$|(f^n(x))'_{x_0}|\neq 1, \tag{24}$$

就称x_0为f的双曲周期点. 对连续可微的单调的函数f，结构稳定的充要条件正是要求所有的周期点都是双曲的.

对于多维流形M上的映射F，如果X_0是它的n-周期点，而且$F^n(X)$在X_0处的Jacobi阵$\left(\frac{\partial F}{\partial X}\right)$不具有模为1的特征值，则称$X_0$为$F$的双曲周期点. 如果$F$是可微分的同胚，则$F$结构稳定的基本

条件仍是周期点的双曲性.

比结构稳定条件稍弱一点的是 Ω 稳定性. 即映射在非游荡集上的稳定性.

非游荡集是微分动力系统中又一个重要概念. 要介绍这个概念,可以先讲什么是游荡点.

如果有 X_0 的一个邻域 $B(X_0)$,在 F 的作用下,$B(X_0)$ 里的点都一去不返 —— 即存在 $N>0$,使对一切 $n>N$,有

$$F^n(B(X_0)) \cap B(X_0) = \varnothing \tag{25}$$

($\varnothing$ 是空集),

就称 X_0 为 F 的游荡点.

游荡点的一去不返,相当于物理过程中昙花一现的"暂态". 对暂态,在研究中可以少考虑,重要的是那些能反复出现的状态,即非游荡点所代表的状态. 映射 F 的所有非游荡点的集合,叫做 F 的 Ω 集.

如果 F 通过 C^r 扰动变成 F^*,而当扰动足够小时 F 与 F^* 作为 Ω 集上的映射总是拓扑共轭的,就说 F 是 Ω 稳定的.

关于一般流形上微分动力系统的 Ω 稳定性,即微分同胚的 Ω 稳定性,也已有深刻的结果. 大体上仍是双曲条件. 如把同胚改为映射,则双曲条件应改为"扩张性". 这方面,我国青年数学家张筑生博士做了很有趣的研究.

对区间或圆周上连续自映射的 Ω 集的性质,Ω 集与周期点集的关系,国内外专家有相当多的探讨. 我国数学家熊金城教授、周作领教授有不少工作. 参看[3].

特别值得提到的,是北京大学廖山涛教授在结构稳定性方面的一系列深刻结果[7]. 关于结构稳定的一般基础知识及近年发展,可参看[8].

§5 马蹄与混沌

为了说明结构稳定的双曲型微分同胚可以复杂到何种程度，Smale 曾构造了一个有名的"马蹄"型微分同胚的例子.

如图 9，映射 T 把矩形 $ABCD$ 变到自己内部，做法是这样的：把矩形沿 AD 方向拉成细长条，但上下两块（即 $ABFE$ 与 $CDHG$ 两个矩形）却缩成两个小帽（图中阴影部分）. 然后把长条弯成马蹄状放在矩形 $ABCD$ 之内，并使两个小帽落在矩形 $ABFE$ 之内，同胚 T 便定义好了.

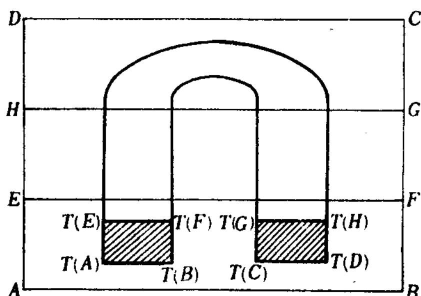


图 9

这个同胚是 Ω 稳定的，而且还可以略加改造使它是结构稳定的. 它具有下列有趣特点：

(1) 它的 Ω 集是二维 Cantor 完全集；

(2) 它具有一切 n- 周期点，而且周期点集在 Ω 集内稠密；

(3) 在 Ω 集中有不可数无穷多这样的点：它在 T 作用之下产生在 Ω 集中稠密的轨道；

(4) 存在正数 d，使得不管 $\varepsilon > 0$ 多么小，对 Ω 集中的任一点 X 总可找到 Y，使得

$$\|X - Y\| < \varepsilon, \text{但} \overline{\lim_{n\to+\infty}} \|T^n(X) - T^n(Y)\| > d,$$

也有这样的 Y：

$$\|X - Y\| > d, \text{但} \lim_{n\to+\infty} \|T^n(X) - T^n(Y)\| < \varepsilon.$$

这里特点(2)、(3)、(4) 描绘出令人眼花缭乱的紊乱现象：在开始时刻十分相近的状态后来可以离得颇远，开始时离得颇远的状态可以变得很近，很多状态在游动，变化，走遍各种不同状态而呈现出周期性状，但又非简单的周期重复. 这种现象被叫做"混沌"(Chaos).

对马蹄的研究使用了"符号动力系统的移位映射"技巧. 这一技巧已变成研究混沌的基本手段. 下面，我们通过一个十分简单但又类似于马蹄的一维映射，说明一下这种技巧.

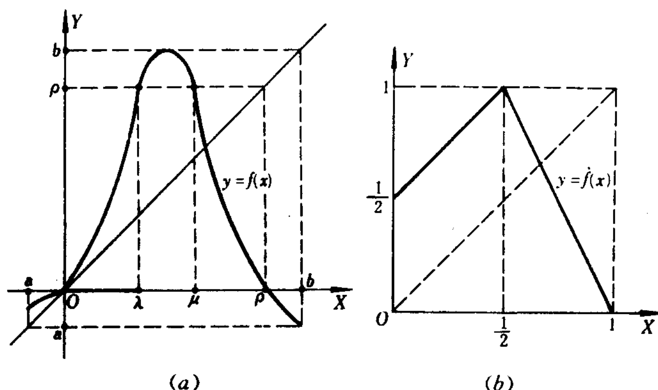


图 10

图 10(a) 可以叫做"一维马蹄". 在[a,b]上给定了一个连续可微的函数 $f(x)$. 把[a,b]如图那样分为 5 个小区间[a,0]，(0、λ)，[λ、μ]，(μ,ρ)，[ρ,b]，顺次记作 $\triangle_0$、$\triangle_1$、$\triangle_2$、$\triangle_3$、$\triangle_4$. 并要求在 $\triangle_1$ 和 $\triangle_3$ 上有 $|f'(x)| \geqslant r > 1$.

对任一点 $x \in [a,b]$，作一个序列

$$I(x) = I_0 I_1 I_2 \cdots I_k \cdots \tag{26}$$

$$(\text{若 } f^k(x) \in \triangle_j, \text{则 } I_k = j),$$

这个序列反映了在 f 的迭代作用之下 x 的象在 5 个小区间内游动的情形. 不妨称它为 x 的"踪迹". 容易看出 $I(x)$ 的如下性质：

1° $I(f(x)) = I_1I_2\cdots I_k\cdots$，一般地：

$I(f^n(x)) = I_nI_{n+1}\cdots I_k\cdots.$

2° 若 $I_k = 0$，则对任意 $l \geqslant 0$ 有 $I_{k+l} = 0$.

3° 若 $I_k = 4$，则 $I_{k+1} = 0, I_{k+l+1} = 0$.

4° 若 $I_k = 2$，则 $I_{k+1} = 4, I_{k+2+l} = 0$.

5° 若 $I_k = 1$ 或 3，则 $I_{k+1} = 1, 3$，或 4.

6° 对于任意一个由 1 与 3 两种数字组成的无穷序列 $I_0I_1\cdots I_k\cdots$，有唯一的一个 x 使 $I(x) = I_0I_1\cdots I_k\cdots$.

7° 设 $I_0I_1\cdots I_m$ 是由 1 与 3 组成的序列，而

$$I(x) = I_0I_1\cdots I_mI_{m+1}I_{m+2}\cdots,$$

$$I(y) = I_0I_1\cdots I_mJ_{m+1}J_{m+2}\cdots,$$

则
$$|x-y| \leqslant \rho r^{-m}. \tag{27}$$

要证明 7° 很简单：由于当 $k = 0, 1, 2\cdots, m$ 时我们有 $f^k(x)$ 与 $f^k(y)$ 在同一个区间 $\triangle_1$ 或 $\triangle_3$ 内，故有 $\triangle_1$ 或 $\triangle_3$ 内的 $\xi_1, \xi_2, \cdots, \xi_m$ 使

$$\begin{aligned}|f^m(x) - f^m(y)| &= |f'(\xi_1)| \cdot |f^{m-1}(x) - f^{m-1}(y)| \\ &= \cdots \\ &= |f'(\xi_1)f'(\xi_2)\cdots f'(\xi_m)||x-y| \\ &\geqslant r^m|x-y|.\end{aligned} \tag{28}$$

再由 $|f^m(x) - f^m(y)| < \rho$ 即得(27). 用(27)及区间套定理可得 6°.

这种利用 $f^k(x)$ 落入的不同区域而把点 x 用一串符号表示的方法，叫"符号动力系"方法. 这里作 f 的迭代时对应的序列只需移动一下各项的位置，因而叫做"移位映射".

从性质 1° — 7° 容易得到混沌性质(2)、(3)、(4). 事实上，由 1 与 3 两种数字构成的序列对应于非游荡点. 其中 n 节循环列对应于 n- 周期点，这表明有各种周期的周期点. 而任一序列都可以和周期列有前 n 项相同，这表明周期点在 Ω 集中稠密. 由于序列中各项的异同刻划出 $f^k(x)$ 与 $f^k(y)$ 是否同在 $\triangle_1$ 或 $\triangle_3$ 中，这就容易

设计出足够多的 y,使混沌性质(3) 与(4) 出现.

如果对图 10(b) 中的函数

$$y(x)=\begin{cases}2x & \left(0\leqslant x\leqslant \frac{1}{2}\right),\\ 2(1-x) & \left(\frac{1}{2}< x\leqslant 1\right)\end{cases} \tag{29}$$

作类似的讨论,可以更容易地证明诸混沌性质. 这时,周期点集在 $[0,1]$ 上稠密,因而 $[0,1]$ 就是 Ω 集. 而 x 的二进小数本身,就可以作为描述 x 的踪迹的一种方法.

迭代本是一种决定性过程. 但由于踪迹序列的任意性和无穷性,使得当迭代次数足够多时 x 究竟落入哪个区间成为随机事件——因为初值的选取不可能绝对精确. 事实上,人们在计算机上作了大量迭代试验. 即使是简单的二次函数 $f(x)=1-ax^2(-1\leqslant x\leqslant 1)$,当 $a>1.5$ 时,$f^n(x)$ 到底是正是负,对多数 x 来说将出现如同掷硬币出现正反面那样的随机现象!

严格的因果关系所支配的决定性系统,竟表现出随机特征. 这与统计物理的研究相映成趣:在统计物理中,在随机性的前提下导出了宏观的决定性因果关系!前者,是时间上的微观决定性导出了宏观上的随机性;后者,是空间上的微观随机性导出了宏观上的决定性. 微分动力系统的研究,从新的角度揭示出决定性与随机性的对立与统一.

对混沌的研究,已成为一个多学科交叉的活跃领域. 数学家,理论物理学家,以及其他一些学科的专家们对它怀有巨大的期望与兴趣[4]. 甚至有人估计,这个领域的进展也许能解开多年来困扰着流体力学家的湍流本质的难题.

微分动力系统是一个迅速发展的研究领域. 本文仅仅涉及一些极粗浅、极不全面的基本知识. 有许多重要而有趣的内容如拓扑熵[6]、遍历性、封闭引理、奇异吸引子等则尚未提及. 有兴趣的读

者可进一步参看所附的文献，以及文献中提供的资料.

参 考 资 料*

1. 张筑生. 微分动力系统. 北京：北京大学出版社，1986.
2. 张景中，熊金城. 函数迭代与一维动力系统，四川教育出版社，成都，1992.
3. 周作领. 一维拓扑动力系统. 数学季刊，1988 (2).
4. 郝柏林. 物理学进展，1983，3，(3).
5. 张景中，杨路. 与 Sarkovskii 序有关的一些定理. 数学进展，1987，16(1).
6. 刘旺金. 关于拓朴熵. 数学进展 11:2，1982，89－100.
7. 廖山涛. 微分动力系统的定性理论. 北京：科学出版社，1992.
8. 张锦炎，钱敏. 微分动力系统导引. 北京：北京大学出版社，1991.
9. 张芷芬，丁同仁，黄文灶，董镇喜. 微分方程定性理论. 北京：科学出版社，1985.

* 这里列出的都是可读性较强的综合性资料，未列出专题论文.

叶彦谦
（南京大学数学系）
Ye Yanqian
(Department of Mathematics,
Nanjing University)

多项式常微系统的定性理论
Qualitative Theory of Polynomial Ordinary Differential Systems

---

中国数学发展的若干主攻方向
Some Main Aspects on Development of Mathematics in China

多项式系统的定性理论是动力系统理论的一个新方向. 三十多年前,中国和苏联数学家在这方面先走了一步. 但近十年来美国和西欧的数学家对此感兴趣的也愈来愈多了. 本文重点介绍二次系统定性理论中的主要成果和研究课题,然后简要介绍它们在 ≥ 3 次多项式系统方面的推广.

§1 历史简述

与微积分发明的同时就有了微分方程(参见[1] 第十章). 早期的数学家致力于求微分方程的通解和研究线性方程的通解结构. 如 L. Euler, G. W. Leibniz, I. Bernoulli, D. Bernoulli, A. C. Clairaut, J. R. D'Alembert, J. L. Lagrange, J. L. Liouville 等,现在仍有以他们的名字命名的可积方程. 但到 19 世纪中叶, Liouville(他还以 Sturm-Liouville 理论而著称) 证明:即使像

$$\frac{dy}{dx} = p(x)y^2 + q(x)y + r(x) \tag{1}$$

(称为 Riccati 方程,其中 p, q, r 是 x 的连续函数) 这样简单的方程,一般来说,也不能用初等积分法来求出通解. 于是数学家转而研究微分方程的基本理论(初值问题的解的存在性和唯一性),微分方程(特别是线性微分方程) 的解析理论(参见[2]) 和微分方程的定性理论. 注意,方程(1) 对于未知函数和它的导数来说是非线性的;而具有两个未知函数的线性方程组却可以通过变量代换化成 (1) 的形状.

定性理论创始于 H. Poincaré 在 1881 − 1886 年间所发表的四篇论文"论微分方程所确定的曲线". 在他的第二篇论文中一些特殊的实多项式常微系统的全局结构分析起到了验证他所首先发现的"极限环"的存在,不存在或唯一性的作用. 例如,他证明二次系统:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = x^2 + y^2 - 1, \frac{\mathrm{d}y}{\mathrm{d}t} = 5(xy - 1)$$

没有有限远奇点，在赤道上有三对（直径相对）奇点，其中两对为结点，一对为鞍点，并画出了相图，即

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{5(xy - 1)}{x^2 + y^2 - 1}$$

在 x,y 平面上的积分曲线族的图形. 为了弄清楚积分线在无穷远处的性态，Poincaré 用球心投影的办法把位于 (x,y,z) 空间中 $z = 1$ 平面上的，上述方程的积分线投影到上（或下）半单位球面 $x^2 + y^2 + z^2 = 1$ 上去，于是赤道 $x^2 + y^2 = 1, z = 0$ 上及其附近积分线的性态就代表原方程的积分线在无穷远处的性态.

注意：在以后出现的方程中，所有的 t 都表示时间参数，而不是空间坐标. 为简单计，我们常用 $\dot{x}, \dot{y}, \cdots$ 代替 $\frac{\mathrm{d}x}{\mathrm{d}t}, \frac{\mathrm{d}y}{\mathrm{d}t}, \cdots$.

他证明三次系统：

$$\dot{x} = x(x^2 + y^2 - 1) - y(x^2 + y^2 - 1),$$

$$\dot{y} = y(x^2 + y^2 - 1) + x(x^2 + y^2 - 1)$$

有唯一的有限远奇点 $(0,0)$，$x^2 + y^2 = 1$ 是极限环，赤道也是极限环. 一切其他的轨线（即 $\frac{\mathrm{d}y}{\mathrm{d}x}$ 的积分线）都是螺线，它的一端绕向 $x^2 + y^2 = 1$，另一端绕向赤道或奇点 $(0,0)$.

$$\begin{cases} \dot{x} = x(x^2 + y^2 - 1)(x^2 + y^2 - 9) - y(x^2 + y^2 - 2x - 8), \\ \dot{y} = y(x^2 + y^2 - 1)(x^2 + y^2 - 9) + x(x^2 + y^2 - 2x - 8) \end{cases}$$

（五次系统）

有极限环 $x^2 + y^2 = 1$，内部有焦点 $(0,0)$. 圆 $x^2 + y^2 = 9$ 是分界线环，其上有结点 $\left(\frac{1}{2}, \frac{1}{2}\sqrt{35}\right)$ 和鞍点 $\left(\frac{1}{2}, -\frac{1}{2}\sqrt{35}\right)$. 赤道上没有奇点，且处处和轨线横截相交.

$$\dot{x} = AC - B, \dot{y} = BC + A, \qquad \text{（七次系统）}$$

其中 $A = x(2x^2 + 2y^2 + 1)$, $B = y(2x^2 + 2y^2 - 1)$, $C = (x^2 +$

$y^2)^2+x^2-y^2+0.1$，有焦点$\left(0,\pm\frac{1}{\sqrt{2}}\right)$和鞍点$(0,0)$. 过$(0,0)$有8字形分界环$(x^2+y^2)^2+x^2-y^2=0$. 在8字形的每一圈内部各有一极限环，分别环绕$\left(0,\frac{1}{\sqrt{2}}\right)$与$\left(0,-\frac{1}{\sqrt{2}}\right)$，它们的方程合在一起可用$C=0$来表示. 赤道处处与轨线相横截.

Poincaré 活到 1912 年，可惜自 1886 年以后的 26 年中他极少在这方面再写文章了. 在他以后微分方程定性理论的发展并不快，读者可以从本世纪五六十年代出版的三本常微专著[3]，[4]，[5]中看出. 在 30 年代以前这方面的著名数学家有 A. M. Liapunov，G. D. Birkhoff，A. Markov，H. Dulac，A. Denjoy，O. Perron，I. Bendixson，M. Frommer 等人. 但 Liapunov 的主要工作是n维系统奇点的稳定性理论，Birkhoff 与 Markov 以抽象动力系统的研究而著称，Denjoy 在环面动力系统方面有很好的工作，Perron 与 Frommer 的许多工作属于n维系统解的渐近性态和平面定常系统的高阶奇点的研究. 所以对平面定常系统的大范围定性理论真正有重要贡献的可以说只有 Dulac 和 Bendixson 两个人. 但是 Bendixson 的工作主要是关于一般平面定常系统的，即现在通常称为 Poincaré-Bendixson 理论的，以及奇点的指标公式；剩下来只有 Dulac 一人了. 他有两项重要工作. 一项是关于有细焦点型或细鞍点型奇点的二次系统的可积条件的研究(1903). 结果很完全，但可惜未总结成一个定理. 另一项是证明任一确定的n次实多项式系统最多只能有有限个极限环(1923)；但证明有漏洞，直到最近几年中才被几位现代数学家所补足. 此外，由他所首创，协助 Bendixson 定理证明不存在闭轨线的，所谓 Dulac 函数法，以后在中国和苏联数学家的工作中被广泛地应用，值得在此一提. 又 M. Frommer 和荷兰数学家 W. Kapteyn 也曾对二次系统存在中心的条件以及全局相图做过研究.

直到1952年苏联数学家 Н. Н. Баутин 关于二次系统的中心或

细焦点经扰动可产生极限环的文章发表以后(摘要发表于 1939 年),多项式系统定性理论才进入一个新的时期. 虽然二次系统的细焦点最多只能是三阶(阶数的定义以后再说)的事实际上从 Dulac 1903 年的论文中已可看出,但 Баутин 的功绩在于首先引起数学家们对研究二次系统的极限环和 Hilbert 第 16 问题后半部的兴趣,并且给出有关这方面的第一个重要结果. 他的由焦点改变稳定性而产生极限环的方法虽然只是现在熟知的所谓 Hopf 分支的一个特例,但使用效果更为显著,所得极限环个数也不只一个,而可能有三个. 1955 年 И. Г. Петровский 与 Е. М. Ландис 发表 40 页的长文试图证明任何二次系统在大范围来说最多也只能有三个极限环. 后来发现不但他们的证明有错误,而且结论也错了(后者是中国人的功绩);但是他们的工作仍然起了推动中国和苏联数学家开展多项式系统实复域定性理论研究的作用.

国内从 1956 年开始,首先有秦元勋,叶彦谦和董金柱的工作. 以后 30 年中国内关于二次系统的工作大量涌现,其中工作成绩较多或较突出的有王明淑,史松龄,马知恩,田景黄,朱德明,任永泰,刘钧,陈兰荪,陈广卿,陈叔平,陈伟峰,李承治,李开泰,李继彬,李孝贵,汪儿年,汪中位,杜星福,余澍祥,沈伯骞,居乃旦,罗定军,杨信安,杨宗培,周孔容,张平光,张芷芬,梁肇军,徐世龙,索光俭,黄启宇,黄克成,盛立人,曹振中,曹玉林,蔡燧林,程中瑗,郭维烈,韩茂安,戴国仁(以姓氏笔画为序)等人.

在前苏联除 Н. Н. Баутин 以外,50 年代以后有 А. Н. Берлинский, Л. С. Лягина, Н. А. Сахарников, Н. А. Лукашевич, Л. А. Черкас, Л. И. Жилевич, Г. С. Рычков, И. С. Куклес, И. Г. Розет, К. S. Сибирский, А. А. Илъин, Ю. С. Илъяшенко, В. Г. Романовский, Н. И. Вулпе 等人. 在美国和西欧自 60 年代以后有 W. S. Loud, L. Markus, L. M. Perko, M. Jacobs, C. Chicone, D. S. Shafer, P. Curtz, J. Argemi, J. Llibre, N. G. Lloyd, D. Koditschek, K. S.

Narendra, A. Lins, J. Guckenheimer, B. A. Coll, A. Gasull R. E. Kooij, C. Christopher, J. W. Reyn, P. de Jager, A. Zegeling, 在加拿大有 D. Schlomiuk, C. Rousseau ,P. Joyal, 在巴西有 R. Bámon, J. Sotomayor,在澳大利亚有 W. A. Coppel ,C. A. Holmes 等. 迄今为止,国内外关于二次系统的论文根据 J. W. Reyn[6] 的统计已在 600 篇以上了. 研究的课题是多方面的,成果也是非常丰富的.

§2 一些最基本的定理和重要研究课题

一、关于奇点

定理 1 实二次系统一般有四个实的有限远初等奇点,三个实的无限远初等奇点;其指标的分布有下列三种可能:

	有限,指标 + 1	有限,指标 − 1	无限,指标 + 1	无限,指标 − 1
Ⅰ	2 个	2 个	2 个	1 个
Ⅱ	3 个	1 个	1 个	2 个
Ⅲ	1 个	3 个	3 个	0 个

注意:有限(或无限)远实奇点有可能重合而成为高阶奇点,或消失而成为复奇点,有限远奇点也有可能跑向赤道而与无限远奇点重合,不一一在此列举.

定理 2 即使二次系统有三个指标 + 1 的有限远奇点,但其中最多只能有两个是焦点型的.

定理 3 当二次系统的四个实初等奇点构成凸四边形的顶点时,两个对顶点有指标 + 1,另两个对顶点有指标 − 1;它们如果构成凹四边形的顶点,则三个外顶点有相同的指标 + 1(或 − 1),一个内顶点则有指标 − 1(或 + 1).

二、轨线与直线的相切或横截性

定理 4　任一直线与二次系统的轨线最多有两个切点(奇点若在此直线上,也作为切点看待). 若切点多于两个,则此直线本身便是轨线. 一般,任一 m 次代数曲线与二次系统的轨线最多有 $m(m+1)$ 个切点.

定理 5　对于具有有限远奇点的非退化二次系统(即方程右边的两个二次多项式没有公因子),必存在一直线 L,它或者是轨线,或者轨线处处从 L 的同一边穿到另一边(可能除了一个奇点以外). 过有限远初等鞍点的分界线在此点的切线便具有上述性质. 过两奇点 S_1 与 S_2 的直线 l 若不是轨线,则轨线在 l 的三段$\overline{\infty S_1}$, $\overline{S_1S_2}$ 与$\overline{S_2\infty}$ 上的穿过方向依次改变. 连接有限远奇点与无限远奇点的直线或为轨线,或为无切直线.

三、闭轨与奇闭轨的性质,相对位置

定理 6　二次系统的闭轨或其上仅含一个奇点的奇闭轨线必为严格的凸闭曲线,它内部只有一个奇点,是焦点或中心.

定理 7　二次系统的闭轨或都环绕同一焦点型奇点(集中分布),或分别环绕两个不同的焦点型奇点. 但不出现 ⚬⚬⚬(三圆呈品字形) 与 ⚬⚬ 型的分布.

定理 8　二次系统的可定向奇闭轨必为凸闭,但可能含整条,半条或一段直线,也可以含一段赤道孤. 通过两鞍点的有限远可定向奇闭轨必含一直线段,它内部的奇点若非中心,则必为粗焦点. 若存在两个这种奇闭轨,则它们有一公共直线段,且内部的奇点都是中心. 存在不能定向的奇闭轨线,其内部有两个焦点或两个中心. 二次系统没有 8 字形的奇闭轨线.

定理 9　已知二次系统的闭轨可以存在(1,0),(1,1),(2,0),(2 ,1),(3,0),(3,1) 等几种分布,其中前两种是严格成立的,后四种可能有模 2 之差. 又二次系统不存在极限环的(2,2) 分布. 有对

称中心的二次系统若存在极限环，则必为严格的(1,1) 分布.

Hilbert 第 16 问题的后半部希望知道 n 次多项式系统的极限环的最多个数以及它们的相对位置. 即使对二次系统这问题也还未完全解决.

定理 10 二次系统不能既存在中心又存在极限环. 当它有一中心和一焦点时，此焦点必定是粗的.

定理 11 如果二次系统除了两个焦点型奇点(且其中至少有一为细焦点) 以外还存在第三个有限远奇点，则其极限环必集中分布.

四、发散量

定理 12 二次系统

$$\dot{x}=P(x,y),\ \dot{y}=Q(x,y) \tag{2}$$

的发散量$\frac{\partial P}{\partial x}+\frac{\partial Q}{\partial y}$若不恒等于常数，则

$$\operatorname{div}(P,Q)=\frac{\partial P}{\partial x}+\frac{\partial Q}{\partial y}=0 \tag{3}$$

的轨迹是一直线 L，(2) 的闭或奇闭轨线必定和 L 相交.

定理 13 若直线(3) 经过一非焦点型奇点，则(2) 的闭轨线必集中分布，若直线(3) 经过两初等奇点，则二次系统(2) 无闭轨线.

定理 14 当二次系统(2) 的方程被化成齐次坐标后，则在$(x,1,z)$ 坐标之下，div = 0 有表达式：

$$[y(\overline{P}_x+\overline{Q}_y)-4\overline{Q}]_{y=1}=0; \tag{4}$$

在$(1,y,z)$ 坐标之下，div = 0 有表达式：

$$[x(\overline{P}_x+\overline{Q}_y)-4\overline{P}]_{x=1}=0; \tag{5}$$

而原来的(3) 式现在可以改写为：

$$[z(\overline{P}_x+\overline{Q}_y)]_{z=1}=0. \tag{6}$$

这里$\overline{P},\overline{Q}$ 是 P,Q 的齐次化多项式.

五、分类问题

定理 15 以(0,0)为指标 +1 的初等奇点的二次系统经仿射变换可以化成下列三种标准型的任何一种：

(i) $$\begin{cases}\dot{x}=-[\lambda x+y+bx^2+(2c+\beta)xy+dy^2],\\ \dot{y}=x+\lambda y+ax^2+(2b+\alpha)xy+cy^2.\end{cases} \tag{7}$$

(ii) $$\dot{x}=-y+\delta x+lx^2+mxy+ny^2,$$
$$\dot{y}=x(1+ax+by). \tag{8}$$

(iii) $$\dot{x}=y+ax^2+bxy,$$
$$\dot{y}=-x+\delta y+nx^2+mxy+ly^2. \tag{9}$$

其中方程(8)又可再分为三类：

$$\dot{x}=-y+\delta x+lx^2+mxy+ny^2, \dot{y}=x. \tag{I}$$

$$\dot{x}=-y+\delta x+lx^2+mxy+ny^2,$$
$$\dot{y}=x(1+ax), a\neq 0. \tag{II}$$

$$\dot{x}=-y+\delta x+lx^2+mxy+ny^2,$$
$$\dot{y}=x(1+ax+by), b\neq 0. \tag{III}$$

以(0,0)为焦点型奇点的二次系统除上述三种标准型(7)，(8)，(9)以外，还可化为：

(iv) $$\begin{cases}\dot{x}=\lambda_1 x-y-\lambda_3 x^2+(2\lambda_2+\lambda_5)xy+\lambda_6 y^2,\\ \dot{y}=x+\lambda_1 y+\lambda_2 x^2+(2\lambda_3+\lambda_4)xy-\lambda_2 y^2.\end{cases} \tag{10}$$

一般的二次系统(2)可以经过仿射变换化为下列五种标准型之一：

$$(1)\ \dot{x}=b_{10}+x^2, \dot{y}=Q(x,y). \tag{11}$$

$$(2)\ \dot{x}=x,\ \dot{y}=Q(x,y). \tag{12}$$

$$(3)\ \dot{x}=b_{11}, \dot{y}=Q(x,y). \tag{13}$$

$$(4)\ \dot{x}=b_{00}+xy, \dot{y}=Q(x,y). \tag{14}$$

$$(5)\ \dot{x}=b_{20}x^2+y,\ \dot{y}=Q(x,y). \tag{15}$$

(7)，(8)，(9)，(10)，(14)，(15)又可经过一定的非线性变换化为 Lienard 方程：

$$\frac{d\xi}{d\tau}=\eta-F(\xi),\frac{d\eta}{d\tau}=-g(\xi). \tag{16}$$

六、中心

定理 16 二次系统(7)出现中心的充要条件是 $\lambda=0$ 再加下列三条件之一：

i) $a+c=b+d=0$,

ii) $\alpha(a+c)=\beta(b+d)$,

$$a\alpha^3-(3b+\alpha)\alpha^2\beta+(3c+\beta)\alpha\beta^2-d\beta^3=0, \tag{17}$$

iii) $\alpha+5(b+d)=\beta+5(a+c)=2(a^2+d^2)+ac+bd=0$.

二次系统(10)出现中心的充要条件是至少成立下列四条件之一：

i) $\lambda_1=\lambda_4=\lambda_5=0$,

ii) $\lambda_1=\lambda_3-\lambda_6=0$,

iii) $\lambda_1=\lambda_2=\lambda_5=0$,

iv) $\lambda_1=\lambda_5=\lambda_4+5\lambda_3-5\lambda_6=\lambda_3\lambda_6-2\lambda_6^2-\lambda_2^2=0$. (18)

二次系统(8)出现中心的充要条件是同时成立下列四个等式：

$\delta=0,\overline{W}_1=m(l+n)-a(b+2l)=0$,

$\overline{W}_2=ma(5a-m)[(l+n)^2(n+b)-a^2(b+2l+n)]=0$,

$\overline{W}_3=ma^2[2a^2+2n^2+nl][(l+n)^2(n+b)-a^2(b+2l+n)]=0$. (19)

当线性部分有纯虚特征根的二次系统被写成复方程：

$$\dot{z}=iz+Az^2+Bz\bar{z}+C\bar{z}^2 \tag{20}$$

(其中 $z=x+iy$)时，出现中心的充要条件是成立下列条件之一：

i) $B=0$; ii) $A=-\frac{1}{2},B=1$;

iii) $I_mA=I_mC=0,B=1$; (21)

iv) $A=2,B=1,|C|=1$.

定理 17 二次系统的一个中心区域不能充满全平面.

七、焦点量和鞍点量

定理 18 二次系统(10)在其焦点(0,0)邻近的螺线的后继函数有表达式:

$$\rho = \rho_o[1 + 2\pi\lambda_1\varphi_1 + \bar{v}_3\varphi_3\rho_o^2 + \bar{v}_5\varphi_5\rho_o^4 + \bar{v}_7\varphi_7\rho_o^6], \tag{22}$$

其中

$$\begin{aligned} \bar{v}_3 &= -\frac{\pi}{4}\lambda_5(\lambda_3 - \lambda_6), \\ \bar{v}_5 &= \frac{\pi}{24}\lambda_2\lambda_4(\lambda_3 - \lambda_6)(\lambda_4 + 5\lambda_3 - 5\lambda_6), \\ \bar{v}_7 &= -\frac{25}{32}\pi\lambda_2\lambda_4(\lambda_3 - \lambda_6)^2(\lambda_3\lambda_6 - 2\lambda_6^2 - \lambda_2^2), \end{aligned} \tag{23}$$

诸 φ_j 是 ρ_o 的幂级数,其系数是诸 λ_i 的整函数. 以 λ_i^* 记满足定理16中条件(18)的 λ_i,则当 λ_i 与 λ_i^* 足够接近,且 $|\rho_o|$ 足够小时这些幂级数收敛,且有 $\varphi_j > \frac{1}{2}$.

定义 1 当 $\lambda_1 \neq 0$ 时称(0,0)为(10)的粗焦点;$\lambda_1 = 0, \bar{v}_3 \neq 0$ 时(0,0)为一阶细焦点;$\lambda_1 = \bar{v}_3 = 0, \bar{v}_5 \neq 0$ 时(0,0)为二阶细焦点;$\lambda_1 = \bar{v}_3 = \bar{v}_5 = 0, \bar{v}_7 \neq 0$ 时(0,0)为三阶细焦点. $\bar{v}_3, \bar{v}_5, \bar{v}_7$ 称为(10)的细焦点(0,0)的三个焦点量. 类似的,$\overline{W}_1, \overline{W}_2, \overline{W}_3$ 称为(8)的细焦点(0,0)的三个焦点量.

定理 19 二次系统的细焦点的最高阶数为3,即当 $\lambda_1 = \bar{v}_3 = \bar{v}_5 = \bar{v}_7 = 0$ 时(0,0)是(10)的中心. (10)的三阶细焦点或满足(18)式条件 i v)的中心可以经过小扰动产生(而且最多只能产生)三个小振幅的单重极限环*.

---

* 若 Γ 是 $\dot{x} = P(x,y), \dot{y} = Q(x,y)$ 的极限环,则当 $\oint_\Gamma (P_x + Q_y)\mathrm{d}t \neq 0$ 时 Γ 称为单重极限环,Γ 的稳定性可由此积分的正或负来判定,见[7],§2.

定理 20 若二次系统(2)存在两个细焦点,则它们都只能是一阶的.若它有一个阶数 ≥ 2 的细焦点,则无细鞍点(即 $P_x + Q_y$ 在此点等于零的鞍点),亦无直线解.

定义 2 解析系统

$$\dot{x} = y + P_2(x,y), \quad \dot{y} = x + Q_2(x,y) \tag{24}$$

(其中 P_2, Q_2 是 x, y 的二次以上的项)在鞍点(0,0)的鞍点量或对偶 Lyapunov 常数 V_k^* 是:当我们能找到一个函数 $F(x,y) = (x^2 - y^2) + F_3(x,y) + \cdots$,使沿(24)的轨线有:

$$\dot{F} = \left.\frac{\mathrm{d}F}{\mathrm{d}t}\right|_{(24)} = V_1^*(x^2 - y^2)^2 + V_2^*(x^2 - y^2)^3 + \cdots. \tag{25}$$

定理 21 以(0,0)为鞍点的实二次系统.

$$\dot{x} = x + a_1x^2 + a_2xy + a_3y^2, \dot{y} = -y + b_1y^2 + b_2xy + b_3x^2 \tag{26}$$

的三个鞍点量为:

$$\begin{aligned} R_1 &= b_1b_2 - a_1a_2, \\ R_2 &= a_2b_3(2a_2 - b_1)(a_2 + 2b_1) - b_2b_3(2b_2 - a_1)(b_2 + 2a_1), \\ R_3 &= (a_1b_1 + a_2b_2 - 5a_3b_3)[a_2b_3(4b_1^2 - a_2^2) - a_3b_2(4a_1^2 - \\ &\quad b_2^2)], \end{aligned} \tag{27}$$

$R_1 = R_2 = R_3 = 0$ 是(26)为可积的充要条件.

以(0,0)为鞍点的实二次系统:

$$\dot{x} = y + Lx^2 + Mxy + Ny^2, \dot{y} = x(1 + Ax + By) \tag{28}$$

的三个鞍点量为:

$$\begin{aligned} V_1^* &= M(N - L) - A(B + 2L), \\ V_2^* &= AM(5A + M)[A^2(2L + B - N)] + (L - N)^2(B - \\ &\quad N)], \\ V_3^* &= A^2M(2A^2 - 2N^2 + LN)[A^2(2L + B - N) + (L - \\ &\quad N)^2(B - N)], \end{aligned} \tag{29}$$

$V_1^* = V_2^* = V_3^* = 0$ 是(28)为可积的充要条件.

定理 22 方程

$$\dot{x} = y - \delta x - lx^2 - mxy - ny^2, \dot{y} = x(1 + ax + by) \quad (30)$$

与方程(8)有相同的奇点，若 O 是(8)的指标 $+1(-1)$ 的初等奇点，则它是(30)的指标 $-1(+1)$ 的初等奇点；若 O 是(8)的细焦点(细鞍点)，则它是(30)的细鞍点(细焦点)．且焦点量的表达式与鞍点量的表达式当 O 为(0,0)时十分相似．

定理 23 若在方程(8)中有 $0 < n < -b, m^2 + 4n(n+b) < 0, \delta = 0$，则 $N\left(0, \frac{1}{n}\right)$ 亦为焦点．此时若 $O(0,0)$ 为三阶细焦点，则 N 与 O 有相同的稳定性．

八、极限环的不存在性和唯一性

定理 24 齐二次系统(不妨设右端无公因子)：

$$\dot{x} = a_{11}x^2 + a_{12}xy + a_{22}y^2, \dot{y} = b_{11}x^2 + b_{12}xy + b_{22}y^2$$

可积分，没有极限环，全局相图只有七种不同的拓朴结构．

定理 25 以(0,0)为星形结点的二次系统(不妨设右端无公因子)：

$$\dot{x} = x + b_0x^2 + b_1xy + b_2y^2, \dot{y} = y + a_0x^2 + a_1xy + a_2y^2$$

没有极限环，全局相图只有 17 种不同的拓朴结构．

定理 26 方程(Ⅰ)当 $\delta = 0$ 时或有中心，或无闭轨线．当 $\delta m(l+n) > 0$ 时无闭轨线；$\delta m(l+n) < 0$ 时若有极限环，必为唯一．对于方程$(\text{Ⅱ})_{a=0}$也有与上面相同的结论．

定理 27 二次系统的三阶细焦点外围不存在闭和奇闭轨线．二阶细焦点外围极限环的唯一性已在很多情况得到证明，但问题尚未完全解决．有细鞍点的二次系统若其第一个鞍点量为零，则或是可积分(这时第二、三个鞍点量亦为零)，或是不存在闭和奇闭轨线．

定理 28 二次系统：

$$\dot{x} = -y + \lambda x + \sum_{i+j=2} b_{ij}x^iy^j, \dot{y} = x + \lambda y + \sum_{i+j=2} a_{ij}x^iy^j$$

若存在绕原点的闭轨线，它必定位于

$$x^2+y^2+\sum_{i+j=2}(a_{ij}x-b_{ij}y)x^iy^j=0$$

的余集中包含(0,0)的那一个连通支内.

九、代数曲线解及其对极限环的影响

定理 29 有一直线解的二次系统其极限环不能多于一个. 若二次系统有一直线解和一细焦点，则无级限环，且细焦点只能是一阶的.

定理 30 有二实直线解的二次系统没有极限环；但即使二次系统有三直线解，它也可能存在中心. 实二次系统有一对复共轭直线解时有可能存在极限环，但最多只有一个.

定理 31 若二次系统以一椭圆(或圆)为极限环，则它别无其他极限环，且此椭圆是单重环. 二次系统若有一双曲线解，则无极限环. 有抛物线解的二次系统有可能存在极限环.

定理 32 二次系统的三次曲线解的闭分支不可能是极限环，但有三次曲线解的二次系统有可能存在其他极限环. 又有重点的三次曲线解的封闭部分有可能成为分界线环.

十、有界二次系统

定理 33 有界二次系统必可经仿射变换化成下列三种形式之一：

1) $\dot{x}=a_{11}x+a_{12}y+y^2,\dot{y}=a_{21}x+a_{22}y-xy+cy^2$，其中 $|c|<2$，其他系数满足下列条件之一：

a) $a_{11}<0$；b) $a_{11}=a_{21}=0$；c) $a_{11}=a_{12}+a_{21}=0,a_{12}\neq 0,ca_{21}+a_{22}<0$.

2) $\dot{x}=a_{11}x,\dot{y}=a_{21}x+a_{22}y+xy$，

其中 $a_{11}<0,a_{22}\leqslant 0$.

3) $\dot{x}=a_{11}x+a_{12}y+y^2,\dot{y}=a_{22}y$，其中 $a_{11}\leqslant 0,a_{22}\leqslant 0,a_{11}+a_{22}<0$；或 $a_{11}<0,a_{22}<0,2a_{22}-a_{11}>0$.

定理 34 有界二次系统也可经仿射变换化成：

$$\dot{x}=-y+\delta x+lx^2+mxy,\dot{y}=x(1+ax+by)$$

的形式，其中诸系数至少满足下列三个条件之一：

i) $(b-l)^2+4ma<0,mb<0$;

ii) $(b-l)^2+4ma<0,b=m+a=0,m(l+m\delta)\leqslant 0$, $m\neq o$;

iii) $bl=ma,b=-l,a+b\delta=m,mb<0$.

定理 35 有界二次系统的有限远奇点最多有三个.有一个或两个有限远奇点的有界系统如果存在极限环，则必为唯一，且是单重环.

下面再列举有关二次系统的一些专题研究，但只列专题和主要研究者，不介绍细节.

1. 奇点的分布，性质和重数(A. H. Берлинский，Г. Я. Латипов，И. С. Кукдес，P. Curtz，J. W. Reyn).

2. 在无限远处的奇点与相图(胡钦训，陆毓麒，A. П. Воробьёв，Г. Р. Латипов，З. Н. Курбанова，J. W. Reyn，任永泰).

3. 二次系统最多只有有限个极限环的实域证明(C. Chicone and D. S. Shafer，R. Bámon，B. K. Романовский，盛立人).

4. 结构稳定无环二次系统的分类(Tavares dos Santos，G.，蔡燧林，史松龄，朱德明，R. E. Kooij).

5. $\dot{x}=x(ax+by+c),\dot{y}=y(a'x+b'y+c')$ 的相图(蔡燧林，J. W. Reyn).

6. 有界系统的相图(R. J. Dickson 与 L. M. Perko，杨信安).

7. 有界系统的极限环(杨信安，W. A. Coppel，J. Llibre 等，韩茂安，张芷芬，李承治，叶彦谦).

8. 弦系统(即没有有限远奇点的)相图的拓扑分类(J. Llibre，A. Gasull，盛立人).

9. 二次系统的奇异环(董金柱，Л. А. Черкас，P. de Jager，沈伯

骞，叶彦谦).

10. 有三阶细焦点的二次系统的相图和分支曲线(秦元勋，史松龄，陈明祥，路纲).

11. 只有一个有限远奇点的二次系统的相图(J. Llibre 等).

12. 有三个鞍点和一个非鞍点的二次系统(A. Zegeling).

13. 方程(Ⅰ)当 $\delta = l = 0$ 时的分支曲线(叶彦谦，马知恩，曹玉林). 当 $\delta = m, l = 0$ 时的分支曲线(曹振中，曹玉林).

14. 方程(Ⅰ)当 $\delta = n = 0$ 或 $\delta = m = 0$ 时的分支曲线与相图(梁肇军，杨信安).

15. 有对称中心与细焦点的二次系统的分支图(董雨滋等).

16. 方程(Ⅱ)当 $\delta = m = 0$ 时的相图(罗定军，孙开浚).

17. 方程(Ⅰ)当 $l = 0$ 时的极限环(叶彦谦，王明淑，李开泰，陈广卿).

18. 方程(Ⅰ)当 $m = 0$ 时的极限环与分支(余澍祥，孙建华，Л. И. Жилевич).

19. 方程(Ⅰ)当 $n = 0$ 时的极限环(居乃旦，Л. И. Жилевич，Л. А. Черкас，沈伯骞，徐世龙与 L. M. Perko).

20. 一般的方程(Ⅰ)的极限环(陈广卿，Л. А. Черкас. Л. И. Жилевич).

21. 有高阶奇点的二次系统的相图与极限环(W. A. Coppel，P. de Jager).

22. 方程(Ⅱ)当 $a = 0$ 时的分界线环(Л. А. Черкас，И. С. Куклес，И. Г. Розет，周孔容，C. A. Holmes).

23. 有二阶细焦点的二次系统的极限环唯一性(韩茂安，陈伟峰).

24. 一般的方程(Ⅱ)的极限环(陈兰荪，蔡燧林，张平光，沈伯骞，陈广卿，刘钧，徐思林，孙建华，刘德明等).

25. 二次系统的代数曲线解(Л. А. Черкас，А. И. Яблонский，

Р. М. Евдокименко, В. Ф. Филиппов, 索光俭, 沈伯骞).

26. 由一系周期解分支出极限环(李继彬,陈孝秋,张芷芬,D. Schlomiuk 等,朱德明).

27. 二次系统的不变量理论(К. С. Сибирский, Н. И. Вулпе, 李继彬,刘一戎).

28. 有对称中心的二次系统的极限环与代数分界环(索光俭).

29. 旋转向量场方法在二次系统定性理论方面的深入应用(叶彦谦,Л. А. Черкас,沈伯骞).

30. 化二次系统为 Abel 方程或极坐标方程来研究极限环问题(W. A. Coppel, J. Llibre, M. Carbonell, A. Gasull).

31. 极限环(2,2)分布的不可能性(叶彦谦,叶惟寅,张平光).

32. 射影平面上的二次系统的相图(梁肇军).

§3 二次系统定性理论研究对一般平面自治系统定性理论的促进以及向 n 次多项式系统方面的推广

在§1中我们已提到过 Dulac 函数法在中国和苏联数学家的研究工作中被广泛地用来证明单连通区域中闭轨线的不存在性. 实际上,它还被用来证明环域中闭轨线的唯一性和某些二次系统极限环的集中分布问题. 类似于此,当 G. F. D. Duff 的旋转向量场理论于 1953 年发表以后,立刻就受到中国数学家的重视和推广;并且在研究二次系统极限环的不存在性,其个数的奇偶性,相对位置,以及分界线环的形成与分支等问题上得到应用. 以上两种方法近年来也被广泛地用来研究生物数学方面多项式或非多项式系统的数学模型.

此外,为了解决某些二次系统极限环的唯一性,唯二性,…,促使中外数学家发现更多的证明极限环唯一性,唯二性,以至唯 n 性

的定理. 可惜的是:我们猜想某些二次系统最多只有两个极限环,但迄今尚无法严格证明.

二次系统定性理论向三次系统,…,以至 n 次系统的推广是意料中的事,而且近十余年来在国内外已有不少成果. 当然,即使是三次系统,§2 中所列举的那些定理有的已不成立,有的情况远比二次系统为复杂而困难. 这主要是由于三次曲线的拓朴分类比二次曲线的分类要麻烦得多,且三次系统奇点的性态(如焦点量)的研究也困难得多. 但是也有一些二次系统方面的定理能够立刻推广到 n 次多项式系统去.

参照 §2,我们看到与定理 1 对应的有:

定理 36 实 n 次系统一般有 n^2 个有限远实初等奇点,$n+1$ 个无限远实初等奇点. 在有限远实奇点中指标 $+1(-1)$ 的最多有$\frac{n(n+1)}{2}$个,指标 $-1(+1)$ 的至少有$\frac{n(n-1)}{2}$个;就是说,不同指标的初等奇点的个数之差最大为 n. 另一方面,不同指标的初等奇点的个数之差最小为 1,当 n 是奇数;为 0,当 n 是偶数.

即使是三次系统,尚无与定理 2 对应的定理. 李继彬等曾举出有 5 个中心和 4 个鞍点的可积三次系统之例. 但当三次系统有 6 个指标 $+1$ 的奇点时,它们之中是否必有一个不是焦点型的,尚未完全证明. 不过,对于多项式系统的奇点分析国内有不少工作,见高维新,李存富,葛渭高,杨世藩,权宏顺等的文章.

与定理 3 对应的有:

定理 37 实 n 次系统若存在$\frac{n(n+1)}{2}$个指标 $+1$ 的初等奇点,$\frac{n(n-1)}{2}$个指标 -1 的初等奇点,则它们有可能构成:

$$(2n-1)-(2n-3)+(2n-5)-\cdots+(-1)^{n-1} \tag{*}$$

型分布. 就是说,有 $2n-1$ 个指标 $+1$ 的奇点处在最外圈,它们一

起构成一个 $2n-1$ 边形(不一定凸)的顶点,此 $2n-1$ 边形把一切其他奇点都包含在其内部(有的可能在边上). 然后有 $2n-3$ 个指标 -1 的奇点构成一个 $2n-3$ 边形的顶点,把余下的所有奇点包含在其内部. ……,依此类推. 最后,有一个指标 $+1$ 或 -1 的奇点处在上述一个套一个的多边形的内部. 也容易做出一个 n 次系统使其不同指标奇点个数之和为 n^2,但个数之差为最小.

注意:即使是三次系统,除了可能出现与(*)对应的 $5-3+1$ 型分布以外,还可能出现 $6-3$ 和 $4-3+2$ 型分布;但可证明不存在 $3-3+3$ 型分布. 至于 $n \geq 4$ 次多项式系统,上述不同类型的分布到底有多少,还未有人去研究过.

有趣的是:叶彦谦在证明三次多项式系统的奇点可以存在 $4-3+2$ 型分布时,把早已由 Poincaré 发现,并在[5]的第七章中得到严格证明,但是在近代定性理论中并未受到重视的下述定理

"**定理 38** 设 G 为平面上的单连通域,它有光滑的边界 L,L 上不存在动力系统 $\dot{x}=p(x,y)$,$\dot{y}=q(x,y)$ 的奇点. 设轨线与 L 相外切的点有 γ 个,相内切的点有 σ 个,则 G 内的奇点指标之和为

$$\sum = 1 + \frac{\sigma-\gamma}{2} \text{"} \qquad (\triangle)$$

推广到 L 非光滑且可能通过奇点的情况去. 然后用它来证明 $4-3+2$ 型分布存在的可能性和 $3-3+3$ 型分布的不存在性. 又推广到 L 可以含轨线段的情况去,并用它证明当奇点有 $5-3+1$ 型分布且最外的 5 个奇点是一凸 5 边形的顶点,则这 5 个奇点不可能都是焦点型奇点.

另一方面,公式($\triangle$)又可推广到平面 n 连通域而成为:

$$\sum = 2-n + \frac{\sigma-\gamma}{2},$$

由此易得亏格为 n 的闭可定向曲面 M 的古典 Poincaré 公式:

M 上动力系统的奇点指标总和 $= \sum = 2-2n$.

和定理 4 对应的，易证：

定理 39 任一直线与实 n 次系统的轨线最多有 n 个切点；若切点多于 n 个，则此直线本身便是轨线. 一般，任一 m 次代数曲线与 n 次系统的轨线最多有 $m(m+n-1)$ 个切点.

与定理 6 对应的有：

定理 40 $2n-1$ 或 $2n$ 次多项式系统的一条单闭轨线在任何方向最多只有 $4n^2-4n+2$ 个极值点.

定理 7 和 8 对三次系统已不成立，与定理 9 对应的，关于极限环的相对位置问题，即使对三次系统，情况也要复杂得多. 李继彬等用对三次 Hamilton 系统作小扰动的方法获得三次系统的极限环的多种形式的相对位置，并证明三次系统极限环的最多个数至少是 11. 定理 10 至 14 对三次系统都不成立.

关于由三次或四次 Hamilton 系统或有高阶奇点的系统经过小扰动以产生极限环和分界线环并在参数平面或空间中画出分支图的工作除李继彬等以外，近年来国外研究得很多，他们也用 Picard-Fuchs 方程以及其他方法，并且还用 Macsyma 等计算机协助计算定积分，焦点量和鞍点量. 其中工作较突出的有周修义，F. Dumortier，R. Roussarie，J. Sotomayor，J. K. Hale，C. Rousseau，P. Joyal，A. Zoladek，以及我国的张芷芬，李承治，王铎等.

对于更一般的 (n,m) 次多项式系统，即

$$\dot{x}=P_n(x,y),\dot{y}=Q_m(x,y)$$

的闭轨线的相对位置问题最近 A. Cima 与 J. Llibre 引进扇(Fan)和巢(Nest)及其阶数的概念，得到扇和巢的个数以及它们的阶的一些估计. 此外，他们还得到一些类似于定理36中关于指标为 $+1(-1)$ 的奇点个数的估计，应用此理论于(1,3)系统，得到极限环的可能分布有下列三种：



这里每一个圈代表属于同一个巢的一切圈，但它们内部还可能包含不同个数的奇点，从而分为几个不同的扇. 对(3,2)型系统，则可能有七种不同形式的巢：



但所有这些图形是否都能被(1,3)或(3,2)型多项式系统来实现，则不得而知.

与定理 15 对应，即使是一般三次系统的分类问题我们也还未见到过. 与定理 16 对应，关于右方不含二次项的三次系统有中心的充要条件已早解决. 对于其他关于三、四次多项式系统在特殊情况下有中心的充要条件，在苏联 Л. А. Черкас И. С. Куклес, А. П. Садовский 等做了许多工作，可在[8]及[9]中找到.

与定理 17 对应可证：$2n-1$ 次多项式系统的一个中心区域可以充满全平面，但 $2n$ 次系统的一个中心区域则不能.

与定理 18 对应的，关于三次系统类似于(22)的表达式还没有. 事实上，三次系统的焦点量的最高阶数迄今未为人所知.

与定理 19 对应，李继彬等证明三次系统的一个细焦点经过小扰动有可能产生 6 个小振幅极限环. 但 N. Llyod 等则能获得 8 个.

最近，A. Zoladek 得到三次系统可以有 11 个小振幅极限环的例子.

对于二次系统情况是非常简单的，定理 20 的前半部，到了三次系统情况也变得非常复杂而难以彻底解决，可以看 Lloyd 和他的学生们的工作.

对 $\geqslant 3$ 次的多项式系统没有和定理 21 至 23 对应的结果.

与定理 24 对应，关于齐三次系统的分类问题有 Cima 与 Llibre 的工作，对一般 n 次齐次多项式系统 J. Argemi 做过一些工作，对 $n \geqslant 3$ 的多项式系统没有和定理 26 — 27 对应的结果. 但定理 25 与

28 可被推广到方程的非线性项为齐 $n(\geqslant 2)$ 次的情况去，结果视 n 为奇数或偶数而有不同.

关于代数曲线解对极限环个数的影响，有：

定理 41（索光俭）： 若 n 次多项式系统有 $\geqslant \frac{(n+2)(n-1)}{2}$ 条直线解，则不存在极限环.

定理 42（黄启宇） 若 Kolmogorov 三次系统

$$\dot{x} = xP_2(x,y), \dot{y} = yQ_2(x,y)$$

有一非退化二次曲线解，且不与坐标轴相切，则不存在极限环（后一条件是沈伯骞等加的）.

关于三次系统的极限环的不存在性，存在性，唯一性和唯二性，国内外的工作不少，都是关于某些特殊三次系统，如第二方程右端有缺项的系统：

$$\dot{x} = y, \dot{y} = Q_3(x,y), \qquad (**)$$

或某些由工程，力学，生物数学，化学反应等方面推导出来的三或四次系统. 此外关于 $\dot{x} = y - F(x), \dot{y} = -x$，当 $F(x)$ 为 $2n+1$ 或 $2n+2$ 次多项式时 A. Lins, W. de Melo 与 C. C. Pugh 猜想，其极限环最多只有 n 个. 但仅在 $F(x)$ 为三次多项式时得到证明，$F(x)$ 为 4 或 5 次多项式时问题也未完全解决. 国内外还有一些研究内部包含三个奇点的极限环的存在唯一性问题的工作，如王现，索光俭，韩茂安等. 关于三次系统有界性的研究陈国维对一般的（* *）做过详细的研究. 关于代数曲线解与极限环和可积性的关系最近有 C. Christopher 与 R. E. Kooij 的工作值得注意.

最后要提一下多项式系统定性理论和代数曲线论以及球面上另一类多项式系统之间的关系. 关于前者可以参考[10]的 § 3. 关于后者，M. I. T. Camacho 于 1981 年引进 R^3 中的齐二次向量场：

$$Q(x) = (Q_1(x), Q_2(x), Q_3(x)), x \in R^3, x = (x_1, x_2, x_3),$$

$Q_i(x)(i=1,2,3)$ 为 x_1, x_2, x_3 的二次齐式，$Q(x)$ 在单位球面 S^2 上

导出切向量场 Q_T：

$$Q_T(x) = Q(x) - (x_1Q_1(x) + x_2Q_2(x) + x_3Q_3(x))x.$$

把 $Q_T(x)$ 分别投影到三个平面 $\Pi_j = \{x \in R^3 | x_j = 1\}, j = 1,2,3$，可得三个平面三次系统：

$W_Q(x) = (x_1Q_3(x) - Q_1(x), x_2Q_3(x) - Q_2(x), 0)$，其中 $x = (x_1, x_2, 1)$，

$R_Q(x) = (0, x_2Q_1(x) - Q_2(x), x_3Q_1(x) - Q_3(x))$，其中 $x = (1, x_2, x_3)$，

$S_Q(x) = (x_1Q_2(x) - Q_1(x), 0, x_3Q_2(x) - Q_3(x))$ 其中 $x = (x_1, 1, x_3)$.

由此可见 (x,y) 平面上的二次系统就是在 $W_Q(x)$ 中令 $Q_3(x) \equiv 0$ 的结果. Camacho 与吴葵光，梁肇军等证明 $W_Q(x)$（特殊的三次系统）的轨线有许多性质与二次系统是一样的. 对于上述三个平面三次系统也可以有类似于定理 14 的表达式，其形式更为对称.

根据以上三节中所写的，可以清楚地看出：多项式系统定性理论是有广阔的发展前途的，它是常微定性理论中一个方兴未艾的新方向.

参 考 文 献

1. В. В. Степанов. 微分方程教程（卜元震译，高等教育出版社，1955）.
2. E. Hille, Ordinary Differential Equations in the Complex Domain. John Wiley & Sons, Inc., 1976.
3. В. В. Немыцкий, В. В. Степанов. 微分方程定性论（王柔怀，童勤谟译. 科学出版社，1954，1959）.
4. E. A. Coddington, N. Levinson. Theory of Ordinary Differential Equations, 1955.

5. P. Hartman. Ordinary Differential Equations. John wiley &. Sons, Inc. :1964.
6. J. W. Reyn. A bibliography of the qualitative theory of quadratic systems of differential equations in the plane. Delft Univ. of Technology, Report 89—71, Delft: 1989,1991.
7. 叶彦谦等. 极限环论. 上海科学技术出版社,1984年(第二版).
8. К. С. Сибирский. Алгебраические инварианты дифференциальных Уравнений и Матриц. Кишинев: ИЗДАТЕЛЬСТВО《ШТИИНЦА》,1976.
9. В. В. Амелькин, Н. А. Лукашевич, А. П. Садовский. Нелинейные колебания в системах второго порядка. Минск: Издательство БГУ ,1982.
10. 叶彦谦. 微分方程定性理论的几个研究课题. 见:常微分方程与控制论. 武汉:华中师范大学出版社,1988. 57—70.

李　忠
（北京大学数学系）
Li Zhong
(Department of Mathematics,
Peking University)

复解析动力系统
Complex Analytic Dynamics

中国数学发展的若干主攻方向
Some Main Aspects on Development
of Mathematics in China

复解析函数的迭代理论的研究,应追溯到P. Fatou与G. Julia的时代. 他们在第一次世界大战期间,分别独立地进行了研究,并为这一理论奠定了基础.

80年代(也即在Fatou与Julia的研究之后60年),这种研究得到了复苏,并引起了广泛的兴趣,其中一个重要的因素是电子计算机的应用. 1980年,B. B. Mandelbrot第一次成功地用电子计算机绘制了映射$z\mapsto z^2+c$在参数平面上的分枝图,后来人们又绘制了许多复迭代的Julia集的图形,这些图形的优美与复杂是出人意外的,一些配有彩色的优美图形简直可以与艺术品相比美(见H. O. Peitgen与P. H. Richlet所著《The Beauty of Fractals——Images of Complex Dynamical Systems》). 根据这些图形人们提出了许多数学上的猜想,进一步刺激了这种研究的发展.

复解析动力系统在近十余年来引起广泛兴趣的另一个因素是由于一批著名数学家的参与使得这一理论有了实质性的突破与进展. 特别应该提出的是D. Sullivan的贡献. 他的关于有理映射的稳定域的最终周期性定理及周期稳定域的分类定理,毫无疑问,是Fatou与Julia之后这种研究的最重要的进展. 此外,A. Douady与J. Hubbard对于二次系统的深刻研究,使我们对于Mandelbrot集有了较多的了解.

复解析动力系统的研究与紊动(Chaos)现象与分形几何(Fractal Geometry)有紧密联系. 解析函数的迭代所形成的Julia集合往往具有紊动现象和Fractal的基本特征. 因此,复动力系统的研究为这些方面的探索提供了新的信息.

本文只是复动力系统的一个一般介绍. 我们不试图罗列这一领域的全部进展,只打算就其基本概念与结果作一概括介绍,并在可能的情况下提到若干正在研究的问题.

§1 从牛顿求根方法说起

大家知道，对于一个实系数的多项式

$$p(x) = a_0x^n + \cdots + a_n \quad (a_0 \neq 0),$$

我们可以用牛顿方法求根，也就是说，从一个给定的初值 x_0 出发，考虑迭代序列

$$x_{n+1} = x_n - p(x_n)/p'(x_n),$$

$$n = 0, 1, 2, \cdots$$

在 $p(x)$ 有实根的情况下，一般说来这个迭代序列收敛于 $p(x)$ 的一个实根.

牛顿方法的一个自然推广便是考察复域上的多项式的求根问题. 早在 1879 年 L. A. Cayley 就认真研究过这个问题，其论文的题目为"The Newton-Fourier imaginary problem"，并将上述牛顿求根方法扩充到复域上的多项式时称为 Newton-Fourier 方法. 他建议研究这样的问题：在复平面上确定这样的区域，使得其中的任意一点作为初值时上述牛顿迭代序列都收敛到多项式的某一个根，他设想一个多项式应该有若干个不同的这样的区域，而这些不同的区域对应于多项式的不同的根.

Cayley 证明了，当多项式是有两个不同根的二次多项式时，由两个根的连线的垂直平分线将复平面分作两个区域，在每个区域中任取一点作为初值，其牛顿迭代序列总是收敛到位于该区域中的根.

当初值取在上述垂直平分线上时，其牛顿迭代序列将没有极限.

Cayley 本人很快就发现，他所提的问题的解答对于二次方程是容易的，但对于三次方程就显得十分困难. 由于他认识到这个问题的困难与复杂，他最终放弃了对它的研究.

尽管如此,Cayley 所提出的问题正是后来 Fatou 与 Julia 发展他们有关有理映射迭代理论的一个重要出发点.

我们不妨把 Cayley 问题扩充为更为一般的形式. 设

$$f(z) = p(z)/q(z)$$

是一个有理表达式,其中 p 与 q 是两个互素的多项式. 设 z_0 为任意给定的一个初始值,考虑 f 迭代的序列:

$$z_1 = f(z_0), z_2 = f(z_1), \cdots, z_{n+1} = f(z_n), \ldots$$

这个迭代序列或集合 $\{z_n : n = 0, 1, \cdots\}$ 被称为 z_0 的前向轨道,或简称为 z_0 的轨道.

我们关心的问题是:初值 z_0 落在怎样的区域内时,其轨道是有极限的,而其收敛性是在区域内局部一致的.

很容易发现,要求整条轨道收敛于一点,这个要求是过分强了. 任何一条周期大于一的循环轨道都不可能满足这样的要求. 因此,我们不妨将要求降低为,轨道的任意一个子序列都有一个局部收敛的子序列. 在这种要求下,所求的区域就是下面要定义的 Fatou 集.

§2 Fatou 集与 Julia 集

为了定义 Fatou 集与 Julia 集,我们需要半纯函数正规族的概念.

设 $\mathscr{F} = \{f_\alpha : U \to \hat{\mathbb{C}}\}_{\alpha \in A}$ 是定义在开集 U 上的一个半纯函数族. 若 $\mathscr{F}$ 中的任意一个序列,都有一个子序列在 U 上局部一致收敛,则称 $\mathscr{F}$ 是一个正规族(normal family).

这里的收敛性是对球面度量而言的.

今后我们记 f^n 为 f 的 n 重复合,即

$$f^n = \underbrace{f\circ \cdots \circ f}_{n}, f^0 = id.$$

在这种记法下，点 z_0 的轨道是$\{f^n(z_0): n=0,1,\cdots\}$.

定义 设 $f: \hat{C}\to\hat{C}$ 是一个有理映射，我们称点 $z_0\in\hat{C}$ 为 f 的一个正规点，如果存在点 z_0 的一个领域 U，使得函数族$\{f^n|U\}$在 U 上是正规族. f 的全体正规点所组成的集合被称为 Fatou 集合，记作 $N(f)$或 N.

Fatou 集合有时被称为稳定集合或正规集合.

显然，由定义不难看出 Fatou 集合总是开集合.

定义 有理映射 $f: \hat{C}\to\hat{C}$ 的非正规点所组成的集合，也即 Fatou 集合之补集 $\hat{C}\backslash N(f)$，被称为 Julia 集合，记作 $J(f)$或 J.

显然，Julia 集合总是闭集合.

因为正规族的概念是一个"函数论"概念，所以单从现在的定义尚不能看清 Fatou 集合或 Julia 集合的动力系统上的意义. 不过很快我们就从它们的基本性质上看出：在 Fatou 集合中的轨道总是"稳定的"，而在 Julia 集合上的轨道是很"紊乱"的.

例 1 设 $f(z)=z^2$，很容易看出，这时 f 的 Fatou 集合是单位圆内部与单位圆外部，而其 Julia 集合是单位圆周.

但是这个例子并不典型，因为其 Julia 集合是很规则的曲线，而这种情况仅是个别的.

例 2　设 $f(z) = z^2 + i$，其 Julia 集合是一条无圈曲线 (denbcite)，如图 1.

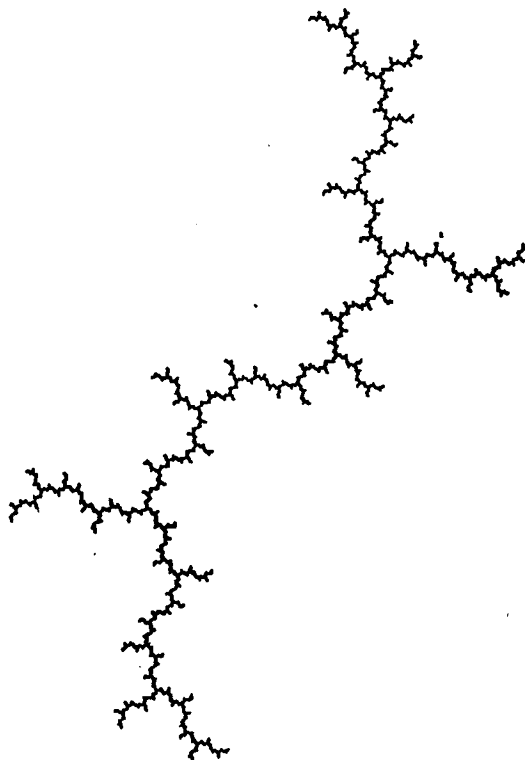


图 1　$z_1 \mapsto z^2 + i$ 的 Julia 集

例3　设 $f(z) = z^2 + (0.99 + 0.14i)z$. 其 Julia 集是一条简单闭曲线,但处处不光滑,如图 2.

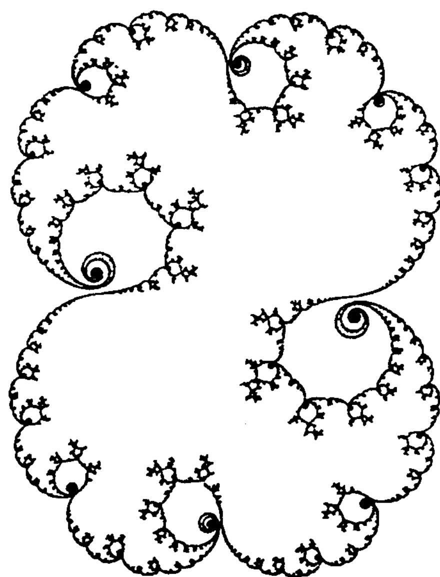


图 2　$z \mapsto z^2 + (0.99 + 0.14i)z$ 的 Julia 集

例 4　设 $f(z)=z^2+c$，其中 $c=-0.765+0.12i$. 其 Julia 集合是一个完全不连通的集合，如图 3.

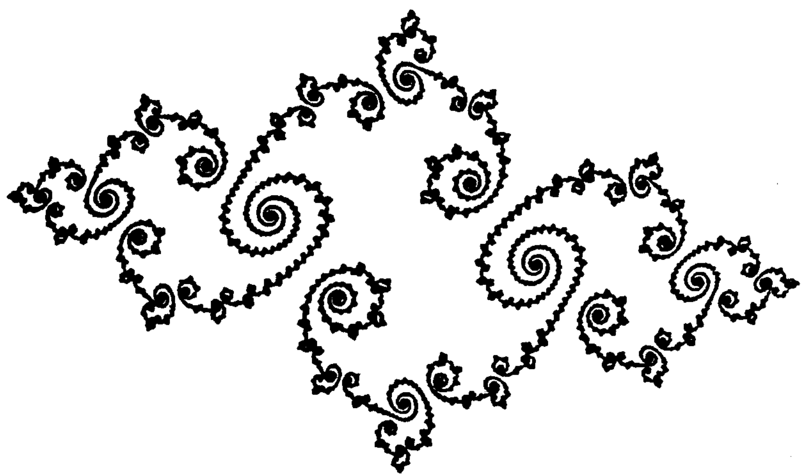


图 3　$z \mapsto z^2+0.12i-0.765$ 的 Julia 集

例 5　设 $f(z)=z^3+\frac{12}{25}z+\frac{116}{125}i$. 其 Julia 集是由无穷多条闭曲线组成. 换句话说，其 Fatou 集由无穷多个连通分支组成，如图 4.

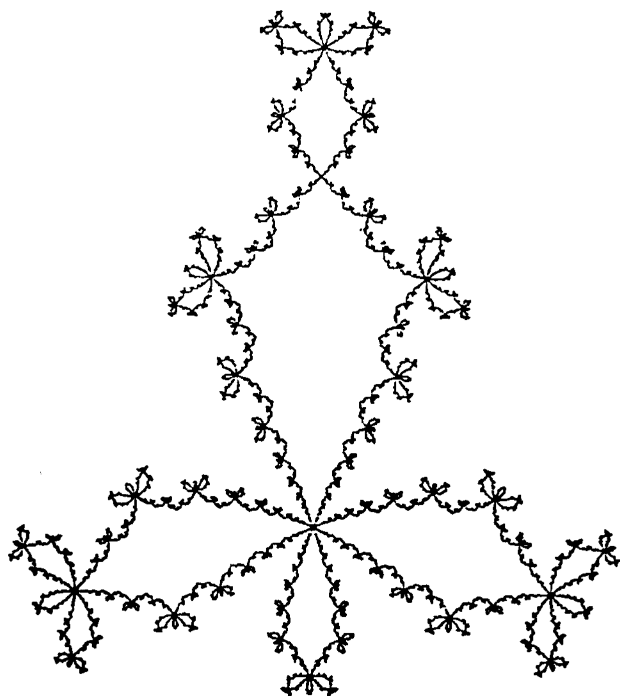


图 4　$z\mapsto z^3+\frac{12}{25}z+\frac{116}{125}i$ 的 Julia 集

以上这些图形都是由电子计算机绘制的. 我们从上述几个例子中已经看到，尽管函数是十分简单的，但其 Julia 集却很复杂. 大多数的 Julia 集合是具有 Fractal 的特征，如处处不光滑以及具有某种自相似性. 但对它的 Hausdorff 维数，我们所知甚少.

我们知道，一个有理映射是复球面到自身的全纯映射；反之亦

然. 因此，对于每个有理映射 $f: \hat{C} \to \hat{C}$ 而言，总有一个正整数 d，使得每个 $w \in \hat{C}$，方程 $f(z) = w$ 恰有 d 个根（将重数计算在内）. 这个数 d 被称为 f 的拓扑度，记作 $\deg(f)$.

很容易验证，当 f 是多项式时，其拓扑度就是多项式的次数；而当 f 表为 $p(z)/q(z)$ 时，其中 p,q 是互素多项式，则有

$$\deg(f) = \max\{\deg p, \deg q\}.$$

因此，$\deg(f) = 1$ 的充要条件是 f 为分式线性变换.

对于分式线性变换而言，其动力系统是平凡的. 例如，双曲型分式线性变换的每条轨道都趋向于其吸性不动点，抛物型分式线性变换的每条轨道都趋向其唯一不动点，而椭圆型分式线性变换共轭于一个旋转变换.

因此，我们只须研究拓扑度大于 1 的有理映射.

命题 1 拓扑度大于 1 的有理映射，其 Julia 集合总是非空的. 换句话说，这时 Fatou 集合不可能是全平面.

Julia 集合有没有可能是全平面呢？回答是肯定的. 第一个这样的例子是由 Lattés 给出.

§3 周期轨道及其附近的轨道的结构

设 $z_0 \in \hat{C}$ 是任意一点，$f: \hat{C} \to \hat{C}$ 是一个有理映射. 若存在一个自然数 n，使得 $f^n(z_0) = z_0$，则称 z_0 为周期点，其轨道被称为周期轨道，或循环. 使得 $f^n(z_0) = z_0$ 的最小自然数 n 被称为周期.

周期轨道在动力系统中常扮演重要角色，在复动力系统中也是如此.

设 z_0 是 f 的一个 n 周期点. 则称数值

$$\lambda = \left[\frac{\mathrm{d}}{\mathrm{d}z} f^n(z)\right]_{z=z_0}$$

为 z_0 的特征值. 根据链规则不难看出

$$\lambda = \prod_{j=1}^{n} f'(f^{j}(z_0)).$$

由此可见，周期轨道上的每一点的特征值都相等. 因此，这个共同的特征值也称为该周期轨道的特征值.

我们根据特征值来对周期轨道进行分类：一个周期轨道被称为

超吸性的 $\Leftrightarrow \lambda = 0$；

吸性的 $\Leftrightarrow 0 < |\lambda| < 1$；

中性的 $\Leftrightarrow |\lambda| = 1$；

斥性的 $\Leftrightarrow |\lambda| > 1$.

对于周期点也有相应的称呼.

Schröder，Koenigs，Leau 及 Böttcher 对于吸性或超吸性周期点作了研究.

Schröder-Koenigs 定理 设 $z_0 \in C$ 是解析函数 f 的一个 n 周期点，其特征值为 λ，且 $0 < |\lambda| < 1$. 则存在 z_0 的一个邻域 U 及共形映射 $\varphi: U \to \triangle_r$（$\triangle r = \{z : |z| < r\}$）使得

$$\varphi \circ f^n \circ \varphi^{-1}(\zeta) = \lambda\zeta,$$

也即下图是可交换的：

$$\begin{array}{ccc} U & \xrightarrow{f^n} & U \\ \varphi \downarrow & & \downarrow \varphi \\ \triangle_r & \xrightarrow[\zeta \mapsto \lambda\zeta]{} & \triangle_r \end{array}$$

Böttcher 研究了超吸性周期点的情况.

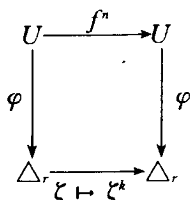
Böttcher 定理 设 $z_0 \in C$ 是解析函数 f 的一个超吸性 n 周期点，且在 z_0 附近有展式：

$$f^n(z) = z_0 + a(z - z_0)^k + \cdots \quad (a \neq 0, k > 1),$$

则存在 z_0 的一个领域 U 及共形映射 $\varphi: U \to \triangle_r$ 使得

$$\varphi\circ f^n\circ\varphi^{-1}(\zeta)=\zeta^k,$$

也即下图是可交换的：



上述两个定理告诉我们，在吸性或超吸性周期轨道附近的点的轨道都被该周期轨道所吸引，并很容易推出

命题 2 吸性或超吸性循环属于 Fatou 集合.

由于在斥性周期点上的特征值的模大于 1，这导致了序列 $\{f^{kn}\mid k=1,2,\cdots\}$ 在该周期点的任意一个邻域内不可能有一致收敛的子序列. 因此，我们证明了

命题 3 斥性循环属于 Julia 集合.

困难的情况发生在中性循环上.

设 z_0 是 f 的一个中性周期点，周期为 n. 如果存在 z_0 的一个邻域 U 及共形映射 $\varphi: U\to\triangle_r$ 使得下式成立

$$\varphi\circ f^n\circ\varphi^{-1}(\zeta)=\lambda\zeta,$$

其中 λ 为 f 在 z_0 的特征值，则容易证明 z_0 属于 Fatou 集合. 反之，若 z_0 属于 f 的 Fatou 集合则上述的邻域 U 及共形映射 φ 一定存在.

上述等式可以看成是关于 φ 的一个泛函方程，通常人们称之为 Schröder 泛函方程，简记为 SFE.

命题 4 中性周期点属于 Fatou 集，当且仅当其相应的 SFE 有解.

设 λ 为某中性周期点之特征值. 若

$$\lambda=e^{2\pi i\alpha},\alpha\in Q,$$

则该中性周期被称为有理中性周期点.

对于有理中性周期点而言，上述 SFE 无解. 事实上，这时存在一个整数 k，使得 $\lambda^k = 1$. 如果 SFE 有解，则必导致

$$\varphi\circ f^{kn}\circ\varphi^{-1}(\zeta) = \zeta,$$

也即

$$f^{kn}(z) = z, \forall z \in U.$$

对于一个拓扑度大于 1 的有理映射而言此为不可能. 因此，我们证明了

命题 5 对于拓扑度大于 1 的有理映射而言，有理中性循环属于 Julia 集合.

是否所有无理中性周期点都在 Fatou 集中呢?Pfiffer 首先给出了这样的例子：α 是无理数，而相应的周期点 z_0 属于 Julia 集.

对于无理中性周期点的一个重要研究工作是属于 Siegel 的(1942 年). 他证明了下面的定理.

Siegel 定理 设 z_0 是一个中性周期点，其特征值 $\lambda = \exp(2\pi i\alpha)$. 若存在两个正实数 a 与 b，使得下式成立

$$\left|\alpha - \frac{m}{n}\right| > \frac{b}{n^a}, \forall m,n \in Z; n > 0,$$

则 z_0 属于 Fatou 集合，也即 SFE 有解.

这个定理的证明是复杂的.

略微粗糙地说，上述条件要求 α 是这样的无理数，在用有理数逼近它时不可能“速度很快”(相对于分母的增大而言).

Siegel 证明了满足上述的条件的值 λ 在单位圆周上是全测度集.

但是，另一方面 Cremer 证明了下面的定理.

Cremer 定理 若无理数 α 满足下列条件：有无穷多个自然数 m 与 n 使得下列不等式成立

$$\left|\alpha - \frac{n}{m}\right| < 1/m^{d^m - 1},$$

则多项式

$$f(z) = z^d + \cdots + e^{2\pi i\alpha}z$$

的中性不动点 $z = 0$ 属于 f 的 Julia 集合.

定义　属于 Julia 集的无理中性周期点被称为 Cremer 点.

尚未解决的问题 1　迄今为止,我们尚不知道,一个无理中性周期点属于 Julia 集与否(即是否是 Cremer 点)这件事只取决于其特征值.换句话说,是否有这样的一个 λ 值,对应于两条周期轨道,一条属于 Julia 集,而另一条属于 Fatou 集?

有理中性周期轨道附近点的轨道的性态已十分清楚.有所谓 **Fatou-Leau 花瓣定理**.其大意是说,每个有理中性周期点都对应若干个吸性花瓣及同样数目的斥性花瓣.落在吸性花瓣内的轨道将被周期点所吸引,而落在斥性花瓣内的点将远离周期点而去.如图 5.

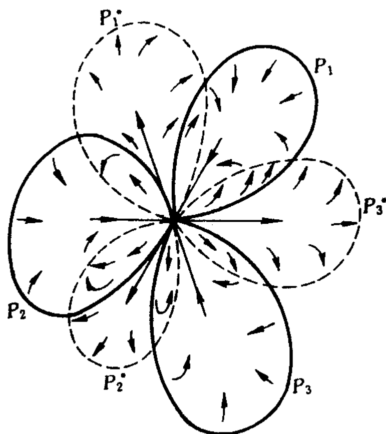


图 5　花瓣定理示意图

尚未解决的问题 2　Cremer 点附近的轨道性态如何?

§ 4 Julia 集的性质

现在我们假定 $f: \hat{C} \to \hat{C}$ 是一个有理映射，其拓扑度 $\deg(f) \geq 2$，用 $J(f)$ 表示其 Julia 集.

命题 7 $J(f)$ 是非空完全集(即没有孤立点的闭集).

由此可见，Julia 集的势为 c.

一个集合 A 称为不变集，如果 $f(A) \subset A$. A 称作完全不变的，如果 A 是不变的，且有 $f^{-1}(A) \subset A$.

命题 8 $J(f)$ 是完全不变的.

由此推出，f 的 Fatou 集也是完全不变的.

由 Julia 集与 Fatou 集的完全不变性可知：一点的轨道，不论是前向轨道还是后向轨道，要么全部属于 Julia 集合，要么全部属于 Fatou 集合.

命题 9 对于任意的自然数 m，f^m 与 f 有相同的 Julia 集，也即

$$J(f^m) = J(f).$$

命题 10 如果 $J(f) \neq \hat{C}$，则 $J(f)$ 没有内点.

从计算机绘制的图形看来，似乎在 $J(f)$ 不是全复球面时，其测度为零. 但是这仅仅是一个猜想.

尚未解决的问题 3 有理映射的 Julia 集合在没有内点时是否其测度为零？即使是对于二次多项式 $p(z) = z^2 + c$，这个问题的一般答案也是未知的.

对于整函数而言，这个问题的回答可以是否定的. 1987 年 McMullen 出人意料地证明了函数

$$f = a\cos z + b$$

的 Julia 集合的测度大于零.

我们已经知道，斥性周期点都在 Julia 集合之内. 下面的命题告诉我们斥性周期点在 Julia 集合上是稠的.

命题 11 $J(f)$ 是 f 的斥性周期点集合的闭包，也即

$$J(f) = \text{closure}\{\text{斥性周期点}\}.$$

我们已知道，$J(f)$ 没有孤立点且非空. 所以命题 11 又进一步推出，拓扑度大于 1 的有理映射有无穷多个斥性周期点.

命题11也可以作为Julia集合的另一个等价定义. 它清楚地告诉我们 Julia 集合在动力系统上的特征.

命题 12 对于任意一点 $z_0 \in J(f)$ 及 z_0 的任意一个开邻域 U，我们总可以找到一个自然 N，使得

$$f^n(J \cap U) = J, \forall\, n \geqslant N.$$

这个性质在动力系统上被称为强混合(strong mixing). 它告诉我们，在 $J(f)$ 上任意取出一个"片段"，对它进行迭代，次数充分大之后，这个片段的像就是 Julia 集的全体.

这是一个很强的性质. 在一定的意义下，它意味着映射 f 在 $J(f)$ 上的作用是紊动的(chaotic).

命题 13 对于任意一点 $z_0 \in J(f)$，我们有

$$J(f) = \text{closure}\{f^{-n}(z_0), n = 0, 1, \cdots\},$$

也就是说，Julia 集合是它的任意一点逆向轨道之闭包；或者说，Julia 集合中的任意一点的逆向轨道在 Julia 集合上都是稠密的.

用动力系统的语言说，f 在其 Julia 集合上的作用是拓扑传递的.

命题 13 是电子计算机绘制 Julia 集的一种理论依据. 就是说，为了画出 Julia 集合，我们只要求出一个它上面的点(比如斥性周期点，或者有理中性周期点)，然后求它的逆像即可.

有理映射在其 Julia 集合上的作用是紊动的，已由上述诸性质看得十分清楚. 首先，它的周期点在 Julia 集合上是稠密的(命题 11). 其次，它有在 Julia 集合上稠密的逆向轨道，从而具有拓扑传递性. 最后，命题 12 告诉我们，它的轨道行为有不可预测性，即对初值具有极端敏感的依赖性.

§5 Fatou 集合的结构

我们已经知道,Fatou 集合是一个完全不变的开集. 但它可能是空集.

利用 Riemann-Hurwitz 公式,很容易证明 Fatou 集 $N(f)\neq\varphi$ 时,只有下列两种可能:

(i)$N(f)$ 有无穷多个连通分支;

(ii)$N(f)$ 仅有一个或两个连通分支.

我们主要关心的问题是在 Fatou 集中轨道行为如何. 如果 $N(f)$ 是第 ii 种情况,那么 f 或 $f\circ f$ 将是 $N(f)$ 的连通分支到自身的映射. 如果是第 i 种情况,这又有两种可能性:

(a)$N(f)$ 的任意一个分支 D 的迭代序列

$$D, f(D), f^2(D), \cdots$$

是最终周期的(eventually periodic),也即存在一个自然数 m,使得 $f^m(D)$ 是周期的.

在这种情况下,f 的某次迭代将 $f^m(D)$ 映为自身,会变得比较容易处理.

(b)$N(f)$ 的某个分支 D 的迭代序列

$$D, f(D), f^2(D), \cdots$$

中任意两个都不相重迭. 这时被称为游荡域.

如果有游荡域发生,问题会变得十分复杂.

Fatou 早在 20 年代就曾猜想:对于有理函数而言,游荡域是不存在的.

这一猜想一直到 1981 年才由 Sullivan 证明.

他的证明用到拟共形映射(quasiconformal mappings)及 Teichmüller 空间理论.

Sullivan 定理 设 f 是拓扑度大于 1 的有理映射. 又设 $D\subset$

$N(f)$ 是 $N(f)$ 的任意分支. 则 D 是最终周期的.

毫无疑问,Sullivan 定理对有理映射的动力系统有重大意义. 它告诉我们,为了弄清在 Fatou 集合上的轨道行为,只要讨论周期稳定域(Fatou 集的连通分支)就足够了.

在 Sullivan 的著名工作发表之后,许多人给出了简化的证明. 在吕以辇的证明中避免了 Teichmüller 理论.

借助于双曲几何,Sullivan 对周期稳定域的轨道行为进行了拓扑分类.

周期稳定域分类定理:设 $D\subset N(f)$ 是有理映射 $f: \hat{\mathbb{C}}\to\hat{\mathbb{C}}$ 的一个稳定域,且 $f^n(D)=D$. 则有下列五种可能性:

(i)D 中有 f 的超吸性周期点,D 中任意一点的轨道都被该超吸性循环所吸引. 在集合

$$\mathscr{A}:=\bigcup_{j=1}^{n} f^j(D)$$

中至少有 f 的一个临界点.

(ii)D 中有 f 的吸性周期点,D 中任意一点的轨道都被该吸性循环所吸引. 在集合 $\mathscr{A}$ 中至少有 f 的一个临界点.

(iii) 在 ∂D 上有 f 的一个有理中性周期点,而 D 中任意一点的轨道均为该有理中性循环所吸引. 在集合 $\mathscr{A}$ 中至少有 f 的一个临界点.

(iv) 在 D 中有一个无理中性周期点 z_0,且 D 是一个单连通域,并存在一个共形映射

$$\varphi: D\to\triangle_r=\{z: |z|<r\}$$

使得 $\varphi(z_0)=0$,且 $\varphi\circ f^n\circ\varphi^{-1}(\zeta)=\lambda\zeta$,$\lambda=\exp(2\pi i\alpha)$,其中 α 为无理数.

(v)D 是一个双连通域,并存在一个共形映射

$$\varphi: D\to\{z: 1<|z|<r\}$$

使得 $\varphi\circ f^n\circ\varphi^{-1}$ 是一个无理旋转.

前三种区域称为 Fatou 域，并依次称为超吸性、吸性和抛物型周期稳定域.

后两种区域称为旋转域，并依次称为Siegel盘和Anold-Herman 环.

我们注意到 Fatou 域对应于至少一个临界点，而拓扑度为 d 的有理映射至多有 $2d-2$ 个临界点. 因此，我们立即得出结论，吸性、超吸性与有理中性循环至多有 $2d-2$ 条.

一个很自然的问题是中性循环至多有多少个？用一个很初等的办法，即将部分中性循环经过一个小扰动之后变成吸性或超吸性的循环，可以证明中性循环至多有 $4d-4$ 条.

这样，非斥性循环至多有 $6d-6$ 条.

很久以来，人们就猜想，非斥性循环至多有 $2d-2$ 条. 这一猜想已为 Shishikura 所证明. 他的证明方法是十分别致的，通过一种拟共形"手术"将所有非斥性循环都变为吸性或超吸性的，而保持映射的拓扑度不变.

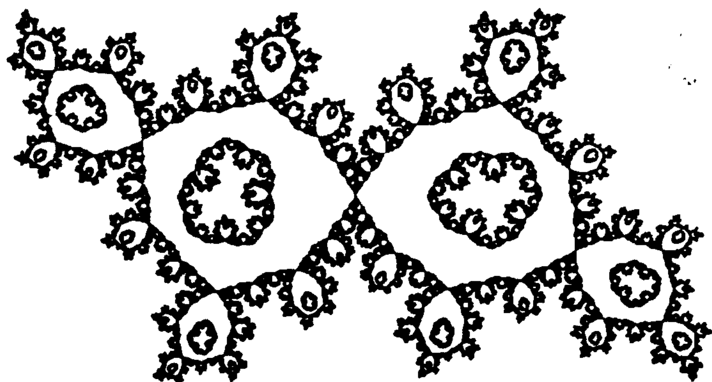


图 6　Anold-Herman 环

Siegel 盘是因 Siegel 在 1942 年的著名工作而得名.

Anold-Herman 环在西方的文献中通常称作 Herman 环，这是1979年 Herman 发现的. 但他在自己的论文中称之为 Anold 环，因为他的论证中用到 Anold 的一定重要定理.

在图 6 中展示了 Herman 环的分布状况.

Fatou 域与临界点有密切的联系. 那么旋转域如何？

我们已经知道，临界点的轨道在旋转域的边界上是稠密的. 但我们不能肯定每个旋转域的边界上是至少有一个临界点. 这是我们提到的第 4 个尚未解决的问题.

定义 一个有理映射的临界轨道都是有穷的，则称为 Thurston 映射.

根据上述讨论，显然Thurston映射没有Siegel盘或Anold-Herman 环.

作为 Sullivan 分类定理的一个推论，我们有

命题 14 若 $f: \hat{C}\to\hat{C}$ 是 Thurston 映射，且每条临界轨道都是预周期的，则 f 的 Julia 集合是整个复球面 $\hat{C}$.

例 设 $f(z)=(z-1)^2/z^2$. 它有两个临界点 $z=1$ 及 $z=0$. 由下述轨道可以看出两个临界点的轨道都是预周期的：

$$\infty\mapsto 1\mapsto 0\to\infty.$$

因此，这个映射的 Julia 集是 $\hat{C}$.

§6 多项式动力系统

设 $p(z)$ 是一个次数大于 1 的多项式. 它区别于一般有理映射的一个重要特征是：∞ 是一个超吸性不动点且 $p^{-1}(\infty)=\infty$. 因此，多项式的 Fatou 集不可能是空集，且其 Julia 集总是 C 上的有界集.

用 $\mathscr{A}_p(\infty)$ 表示 p 的 ∞ 点的吸性盆，即

$$\mathscr{A}_p(\infty):=\{z: p^n(z)\to\infty\},$$

则我们有

命题 15 $J(p) = \partial \mathscr{A}_p(\infty)$.

对于多项式而言,无穷远点是一临界点.其他临界点被称为有穷临界点.有穷临界点的轨道对于 Julia 集的性质有很大影响:

命题 16 若每个有穷临界点的轨道都不趋于 ∞,则 $J(p)$ 是连通的.否则,$J(p)$ 是完全不连通的(即 Cantor 集).

例 设 $p_c(z) = z^2 + c$.它的有穷临界点只有点 $z = 0$.于是,它的有穷临界点的轨道是

$$0, c, c^2 + c, (c^2 + c)^2 + c, \cdots.$$

这样,p_c 的 Julia 集是否连通,就依赖于上述序列是否趋于无穷.

著名的 Mandelbrot 集就是定义为参数 c 平面上的这样的 c 的集合:其中每个 c 使得 p_c 的 Julia 集合是连通的.

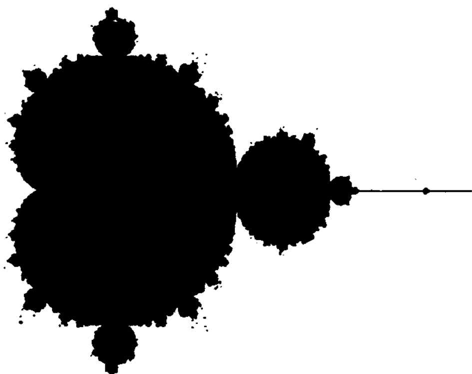


图 7 Mandelbrot 集

如果我们将上述关于 c 的序列的第 n 项记为 $a_n(c)$,那么 Mandelbrot 集也可等价地定义为

$$M := \{c \in \mathbf{C} : a_n(c) \nrightarrow \infty\}.$$

1980 年 Mandelbrot 首先画出了这个集合的图形(如图 7).

对于任何一人说来，从一个具体值 c 出发计算其平方后再加 c，然后将所得值平方再加上 c，如此下去，这是一个简单的运算. "Everyone can do that. But no one expected to find so much enigmatic beauty hidden in this iteration"(见《The Beauty of Fractals》,p. 9). Mandelbrot 集的边界的复杂性及其优美可参见《The Beauty of Fractals》中第 78 — 86 页中的彩色图形.

根据这些图形，人们作了种种猜想. 首先，大家猜想 Mandelbrot 集合是连通的. 其次，由于其边界的极端复杂，人们猜想 ∂M 的 Hausdorff 维数可能为 2. 前一个猜想已为 Douady 与 Hubbad 所证实.

Douady-Hubbard 定理：Mandelbrot 集是连通的.

尚未解决的问题 5 确定 Mandelbrot 集的边界的 Hausdorff 维数.

在 Mandelbrot 集的连通性被证实之后，下述的问题一直是大家十分关注的：

尚未解决的问题 6 Mandelbrot 集是否局部连通的？也就是说，对于任意的开集 $U \subset \mathbf{C}$ 是否总存在一个开集 $V \subset U$ 使得 $V \cap M$ 是连通的？

一般说来，人们倾向于答案是肯定的.

跟 Mandelbrot 集有关的猜想还有一个较为重要，即所谓双曲猜想. p_c 被称为双曲型的，如果它有一个有穷的吸性循环.

尚未解决的问题 7(双曲猜想) Mandelbrot 集的内部全部是由这样的 c 组成，它使得 p_c 是双曲型的.

Mandelbrot 集的局部连通蕴含着双曲猜想.

跟 Mandelbrot 集有关的还有 Julia 集合的局部连通性问题.

尚未解决的问题 8 当 $c \in M$ 时 $J(p_c)$ 是否总是局部连通的？或者说，怎样的 c 才使得 $J(p_c)$ 是局部连通的？

§7 超越整函数的动力系统

设 $f: C \to C$ 是一个整函数，那么，同样可以考虑其迭代及定义其 Julia 集与 Fatou 集. 当 f 是多项式时，前面已作了讨论，因此，只消讨论 f 是超越整函数的情形即可.

最早的一篇有关超越整函数迭代理论的论文是 Fatou 在 1926 年写的. 他在其中证明了 Julia 集的基本性质如同有理函数一样. 但是，Julia 集是斥性周期点集合的闭包这一命题除外，而这一命题一直到 1986 年由 Baker 证明. 他的证明是目前唯一已知的证明，其中用到了 Ahlfors 的一个深刻的定理.

虽然，在基本性质上，整函数的 Julia 集与有理函数的 Julia 集是一样的，但进一步的研究表明，在许多深入一些的性质上两者却有本质的差别.

早在 Sullivan 证明有理函数无游荡域之前，Baker 就举例证明了整函数可能有游荡的稳定域.

因此，一个重要的问题是：哪些超越整函数像有理函数一样没有游荡的稳定域？

为了叙述这方面的结果，我们介绍一些记号与定义.

一个值 $c \in C$ 被称为整函数 f 的渐近值，如果存在一条曲线 τ 伸向 ∞ 且使得 f 沿 Γ 趋向 ∞ 时，其函数值趋于 c.

一个值 $c \in C$ 被称为 f 的临界值，如果它是某临界点所对应的函数值.

函数 f 的全体临界值与渐近值所组成的集合称为 f^{-1} 的奇异值集合，记作 $\mathrm{sing}(f^{-1})$.

一个整函数 f 被称为属于 Σ 类，如果它的奇异值集合 $\mathrm{sing}(f^{-1})$ 是有限的. 通常大家熟知的初等函数 $\exp(z)$，$\cos(z)$ 与 $\sin(z)$ 都是 Σ 类函数. 一般说来，若 g 与 h 是多项式，则

$$f = \int g(z)\exp h(z)\mathrm{d}z$$

属于 Σ 类.

下边的命题实际上是由许多人先后证明的，如 Baker，Eremenko，Lyubich，Goldberg，Keen.

命题 16 Σ 类函数的 Fatou 集的任意连通分支都是最终周期的.

另一类函数称为 B 类整函数，其定义为

$$B:\{f:\text{sing}(f^{-1})\ \text{有界}\}.$$

显然，$\Sigma \subset B$，并且容易证明 $B\backslash\Sigma \neq \varphi$，如

$$(\sin z)/z \in B\backslash\Sigma.$$

B 类函数的一个重要拓扑特征是，存在一个充分大的圆 A，使得

$$f: \mathbf{C}\backslash f^{-1}(A) \to \mathbf{C}\backslash A$$

是一个无歧覆盖.

尚未解决的问题 9 B 类函数是否有游荡的稳定域?

迄今为止，我们所知道的是，如果 B 类函数有游荡域，也只能在一个有界的范围内进行.

整函数的动力系统与有理函数的动力系统的区别还在于周期稳定域之不同. 对于超越整函数而言不可能有 Anold-Herman 环. 对整函数而言，若 Fatou 集的一个分支是多连通域，则必为游荡的，从而不可能是周期的.

对整函数说来，除了 Fatou 域与 Siegel 盘之外，尚有一种周期稳定域，被称为 Baker 域：其中每条轨道都收敛于 ∞.

命题 17 任意一个整函数的周期稳定域只能是下列三种域之一：Fatou 域，Siegel 盘与 Baker 域.

Σ 类函数乃至 B 类函数无 Baker 域.

§8　指数函数族

像在多项式情况下研究二次函数 z^2+c 一样，在超越整函数情况下人们集中研究了指数函数族：$z \to \lambda\exp(z)$，记为 E_λ.

Fatou 曾猜想 $E_1=\exp(z)$ 的 Julia 集合是全平面. 这一猜想于 1981 年为 Misiurewicz 所证实. 后来，Devaney 证明了：当 $\lambda>e^{-1}$ 时 E_λ 的 Julia 集合为全平面.

我们考虑集合

$$\mathscr{D}=\{\lambda\in C: J(E_\lambda)=C\}.$$

这个集合是 Mandelbrot 集合的类比物，但是迄今为止，我们对它所知甚少.

Devaney 用电子计算机画出了它的大致图形（如图 8）. 但是无

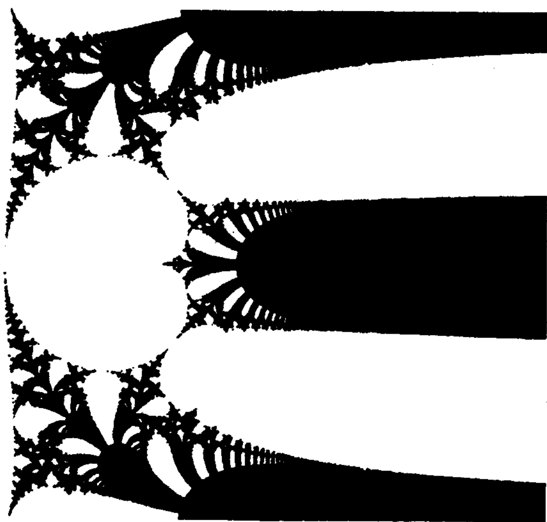


图 8　指数函数族分枝图

法断定这个图形真实地代表着它. 比如,从这张图上看上去是一个区域,但是到目前为止,我们尚不知 $\mathscr{D}$ 中有否内点.

尚未解决的问题9 $\mathscr{D}$ 中是否有内点?$\mathscr{D}$ 的测度是否是正的?

从种种迹象看来,似乎 $\mathscr{D}$ 中没有内点. 在周建莹与作者的论文中证明了 $\mathscr{D}$ 在实轴上没有内点. 这就是说,对于任意实数 $\lambda^* > e^{-1}$,存在 $\lambda_n \in C$, 使得 $\lambda_n \to \lambda^*$ 且 $J(E_{\lambda_n}) \neq C$. 这样,当 λ_n 趋向于 λ^* 时,Julia 集合从曲线族突然变成全平面. 这种现象被称为 Julia 集"爆炸".

§9 其他问题

以上我们考虑了 $\hat{C}$ 到自身与 C 到自身的全纯映射的迭代,很自然地会问,若 $f:S \to S$ 是一般 Riemann 曲面 S 到自身的全纯映射,那情形会如何?

这种推广不可能走得很远,事实上,只有当 S 是下列三种典型 Riemann 曲面时全纯动力系统才是非平凡的:

(i)$S = \hat{C}$,

(ii)$S = C$,

(iii)$S = C\setminus\{0\}$.

根据 Riemann 曲面的单值化定理,除了上述三种曲面之外,其余 Riemann 曲面要么是双曲型 Riemann 曲面,要么是环面,在前一种情形时,迭化的轨道只有几种简单的可能,或者全部是周期的,或者全部为一个循环所吸引,或者全部在 S 上没有极限点. 当 S 是环面时,全纯映射 f 提升到 S 的万有覆盖上是一个仿射变换,其系数还满足特定的方程,值得研究的问题不多.

这样看来,对于一维复流形上的全纯映射只有上述三种情况值得研究. 目前的研究大多集中在(i) 与(ii) 两种情况上. 对于情况(iii) 的研究也是有趣的,但是有些基本性质可通过提升归结到

情况(ii). 在这方面方丽萍同志的文章作了系统讨论.

注　这篇综述文章完成以后,这一领域又有了重要进展,许多新进展及新文献尚未包含进去. 例如,上面提到的第 5 个尚未解决的问题已为 Shishikura 所解决. 另外,国内也有许多很好的研究工作,如吕以辇教授与任福尧教授以及他们的学生的工作. 这些工作均未能提及. 谨此说明.

参 考 文 献

1. L. Ahlfors, Complex analysis, McGraw-Hill, 1979.
2. L. Ahlfors, Lecture on quasiconformal mappings, Van Nostrand, 1966.
3. H. Baker, An entire function which has wandering domains, J. Austral. Math. Soc. (A) 22(1976), 173—176.
4. H. Baker, Wandering domains in the iteration of entire functions, Proc. London Math Soc. 49 (1984), 563—576.
5. M. Bar, Bull. Kasan Math. Soc. 14(1905), 176.
6. H. Cremer, Zum Zentrumproblem, Math. Ann. 98 (1928), 151—163.
7. J. Curry, L. Garnet and D. Sullivan, On the iteration of a rational function computer experiments with Newton's method, Comm. Math. Phys. 91(1983), 267—277.
8. R. L. Devaney, Structural instability of exp (z), Proc. Amer. 94 (1985), 545—548.
9. R. L. Devaney and M. Krych, Dynamics of exp (z), Ergodic Theory Dynamical Systems 4(1984), 35—52.
10. P. Fatou, Sur les equations functionnelles, Bull. Soc. Math. France, 47 (1919), 161—271; 48(1920), 33—94; 48(1920), 208—314.
11. P. Fatou, Sur l'teration der fonctions transcendantes, Acta Math. 47 (1926), 337—370.
12. M. Herman, Exemples de fractions rationelles ayant une orbite dense

sur ta sphere Riemann (preprint).

13. G. Julis, Memore sur l'iteration des fonctions rationnelles, J. Math. 8 (1918), 47—245.

14. S. Lattes, Sur l'iteration des substitutions rationnelles et les function de Poincaré, C. R. Acad. Sci. Paris, 166(1918), 26—28.

15. B. Mandelbrot, The fractal geometry of nature, Freeman, 1982.

16. B. Mandelbrot, Fractal aspects of iteration of $z \to \lambda z(1-z)$ and z, Ann. New York Acad. Sci. 357(1980), 249—259.

17. M. Misiurewicz, On iteration of exp(z), Ergodic Th. Dym. Syst. 1 (1981), 103—106.

18. A. E. Eremenko and M. Yu. Lyubich, Iterates of entire functions, Soviet Math. Dokl. 30(1984), 592—594.

19. L. R. Goldberg and L. Keen, A finiteness theorem for a dynamical class of entire functions, 6(1986), 183—192.

20. C. McMullen, Aera and Hausdorff dimension of Julia sets of entire functions, trans. Amer. Math. Soc. 300(1987), 325—342.

21. D. Sullivan, Quasiconformal homeomorphisms and dynamic[illegible] Ann. Math. 122(1985), 402—418.

22. J. Zhou and Z. Li, Structural instability of mapping $z \mapsto \lambda\exp(z)$ $(\lambda > e^{-1})$, Science in China, A., 32(1989) 1153—1163.

IV

偏微分方程数值解法及在科学、工程中的应用

Numerical Solution of Partial Differential Equations and It's Applications to Science and Engineering

徐福臻
(中国科学院计算中心)
Xu Fuzhen
(Computing Center,
Academia Sinica)

第三种科学方法与计算数学
The Third Scientific Method and Computational Mathematics

摘　要

本文回顾了近代科学发展的历史，介绍了自牛顿起四百年来人类所积累的一些主要的精确的科学技术知识，探讨了牛顿方程、欧拉方程、纳维-斯托克斯方程、麦克斯韦方程、波尔兹曼方程、薛定谔方程等及其派生方程转化为计算机程序并用于实际的意义.

阐述了电子计算机正在把计算的方法推向人类科学活动的前沿，计算与实验、理论相并列，已经成为第三种科学方法.

介绍了“偏微分方程数值解法及其在科学、工程中的应用”一题的创立背景、研究思想、研究方法及其应用.

§1　近代科学的诞生

近代科学开始于西欧文艺复兴时代. 文艺复兴运动发源于意大利，到了15世纪后半叶扩展到整个欧洲，16世纪形成了高潮. “文艺复兴”这个术语来源于法兰西语“Renaissanse”，其含意是指古代文化的复兴. 当时运动的先驱者借助于古代希腊、罗马的古典文化反对封建神学，这就给人们造成了错觉，仿佛这场运动是古典文化的复兴. 其实，不是古典文化的复兴，而是一场资产阶级新文化运动.

当时的封建社会制度使欧洲文化走向了低潮，因为作为封建制度支柱的天主教垄断了整个社会的知识教育，僧侣是唯一的知识阶层，而他们大都愚昧无知，主张大地是一个圆盘，周围有海洋环绕，中心在他们的圣地耶路撒冷；天是由四根大柱支撑的穹窿. 而新兴的资产阶级为了发展工商业，需要了解原料性能，革新工具，通晓生产技术，开辟市场，改进运输，等等，总之急需技术进步，迫切地要把科学从教会的桎梏下解放出来. 于是，伟大的发明和发

现就接二连三地跟着来了. 这是文艺复兴时代科学革命的生产背景.

文艺复兴的思想启蒙运动为科学从神学的束缚下解放出来提供了精神武器. 文艺复兴的主题是用人文主义世界观反对宗教神学世界观. 人文主义世界观要求以人为中心而不是以神为中心观察一切,要求人们不要按照经院哲学家编造的谎言和臆测去认识自然,反对盲从和偶像崇拜. 这是一个锐利的思想武器,人们正是用它来反对神学,建立起近代科学的.

在文艺复兴前后,中国的罗盘针、火药、印刷术、造纸术四大发明,阿拉伯的天文、数学、化学、医学等科学文化,古希腊的自然哲学相继传入了欧洲,为科学革命奠定了科学技术基础.

文艺复兴时代的科学革命的主要内容有两项,一项是太阳中心说,一项是解剖学. 波兰天文学家哥白尼(N. Copernicus,1473—1543)对天体进行了长期观察和研究,发现地球自转,而且和其他行星一起围绕太阳运转. 他于1543年发表《天体运行论》中提出了太阳中心说,用以否定地球中心说,推翻了托勒密体系,这是对宗教神学的第一个决定性打击. 他创立太阳中心说是十分不易的,既要有自然哲学的确实论据,又要敢于冒宗教迫害和传统观念打击的风险. 尽管从天体观察中发现了托勒密的错误,但地球中心说有着自己的权威性,人们日常生活中的直觉观念也认为地球中心说是对的,如太阳总是从东方升起西方落下,上抛的物体仍落回原处,等等. 哥白尼对问题的突破当然主要靠亲自观察,通过三十多年的天文观察;另一方面也借鉴了前人的科学实践与论述,他曾仔细研究过古希腊自然哲学,并从中找到了论据;还有,应用几何学等数学工具加以推演论证. 同哥白尼的《天体运行论》有着同等意义的《论人体构造》是比利时著名医生维萨里(A. Vesalius,1511—1564)的伟大著作. 它以血液循环"以心脏为中心"学说推翻了盖伦的"以肝脏为中心"的理论,揭开了封建愚昧的蒙纱.

人们都把1543年作为近代科学诞生之年，这是因为在这一年《天体运行论》与《论人体构造》这两部具有重大历史意义的著作问世，这两部著作分别奠定了近代天文学与近代医学的基础，成为中世纪科学与近代科学疆界线上的丰碑，它们是近代科学诞生的主要标志. 哥白尼和维萨里是近代科学的先驱.

继哥白尼之后，德国天文学家开普勒(J. Kepler，1571 — 1630)在分析第谷·布拉赫的观测资料的基础上，发现了行星沿椭圆轨道运行，提出了行星运动三定律；意大利物理学家、天文学家伽利略(Galileo，1564 — 1642)发现了自由落体定律，推翻了亚里士多德的物体下落的速度与物体重量成比例的臆测，他还发现了物体的惯性定律，给出了力和加速度的概念，成为后来牛顿建立经典力学的基础. 不过，哥白尼、开普勒、伽利略他们发现的物理概念和规律都还是孤立的.

§2 科学基本规律的精确数学描述

英国伟大的物理学家、数学家牛顿(I. Newton，1642 — 1727)对天体中行星的运行及地面上的物体运动作了整体统一的考察，并采用数学方法表述物体运动过程中的因果关系. 他在1687年发表的《自然哲学的数学原理》一书中提出了力学的三大定律和万有引力定律，把地球上物体的力学和天体力学统一于自然界机械运动形态上，正确地反映了宏观世界中物体低速机械运动的客观规律性，从而建立了成为经典力学基础的牛顿力学体系，也就是我们今天所说的牛顿经典力学. 牛顿的第二定律，即运动定律的数学描述如下：

$$
\left.
\begin{aligned}
F_x &= m \cdot \frac{\mathrm{d}^2 x}{\mathrm{d}t^2}, \\
F_y &= m \cdot \frac{\mathrm{d}^2 y}{\mathrm{d}t^2}, \\
F_z &= m \cdot \frac{\mathrm{d}^2 z}{\mathrm{d}t^2}.
\end{aligned}
\right\}
$$

式中(F_x, F_y, F_z)是作用力矢量在直角坐标系中的分量;m是质点的质量,通常为常量;(x, y, z)是质点坐标.这是质点动力学的基本方程,称为牛顿方程.利用牛顿方程,人们就可以根据物体的运动坐标及速度的初始值,精确地求出该物体的未来或过去的运动状态.

牛顿力学在物理学其他分支,以及其他自然科学中,特别是技术科学中获得了非常广泛而成功的应用.但是,任何理论的正确性都是相对的,需要在实践中不断检验并求得发展.从19世纪到20世纪,物理学家们在实验中发现高速运动着的物体,其质量变化很大.例如,当电子的速度达到0.9光速时,它的质量是静止时的质量的2.3倍.这个事实说明牛顿力学中的"质量不变"是不适用于接近光速运动的物体.于是,产生了相对论力学.这是20世纪初一项重大科学发现.相对论是由著名物理学家爱因斯坦(A. Einstein,1879 — 1955)创立的.爱因斯坦生于德国,1933年迁居美国.他在前人许多新发现的推动下,建立了狭义相对论(1905年);并在此基础上推广为广义相对论(1916年).

在相对论力学中,还认为质量随着能量的增加而增加,反之,能量也随着质量的增加而增加.这一规律在原子核反应中得到了证实.牛顿力学认为物体的质量和它的能量无关,也不是普遍真理.

在相对论中建立了新的时空观和可与光速比拟的高速物体运动规律.但在一般情况下,相对论效应极其微小.因此,牛顿力学可认为是相对论力学在低速情况下的近似.

物理学家还在微观世界中发现了极微小物体的位置和速度不能同时量度得很准确.例如对原子内部运动着的电子来说,测定其速度的误差可达几倍之多.牛顿力学认为物体的位置与速度可同时被确定的规律,在微观世界中也不适用了.于是,又产生了量子力学.量子力学是研究微观粒子运动规律的理论.奥地利物理学家薛定谔(E. Schrödinger,1887 — 1961)是这一理论的奠基人之一.他在前人的物质波理论基础上,于1926年提出了物质波运动方程

$$\frac{h^2}{2m}\left(\frac{\partial^2\varphi}{\partial x^2}+\frac{\partial^2\varphi}{\partial y^2}+\frac{\partial^2\varphi}{\partial z^2}\right)-U(x,y,z)\varphi+ih\frac{\partial\varphi}{\partial t}=0,$$

式中 $\varphi(x,y,z,t)$ 为波函数;$U(x,y,z)$ 为该粒子在外场的势能;m 为质量;h 为普朗克常数.这个方程通常称为薛定谔方程.它是量子力学中描述微观粒子(如电子)运动状态的基本定律,在粒子运动速度远小于光速的条件下适用.量子力学将宏观认识引伸到微观世界.这一基本定律,在量子力学中的地位与牛顿运动定律在经典力学中的地位很相似.量子力学的规律用于宏观物体或质量、能量相当大的粒子时也能得出牛顿力学结论,它是牛顿力学的补充.

这就是说,在特殊情况下必须应用相对论力学或量子力学规律解决问题;在通常情况下应用相对论力学和量子力学的规律就不必要了,应用牛顿力学就可以解决问题.所谓通常情况也包括现代高技术领域中研究的物体运动,例如火箭飞行,它的速度即使达到第二宇宙速度11.2千米/秒,与光速还差4个数量级,牛顿力学对它仍适用.

英国物理学家麦克斯韦(J. C. Maxwell,1831 — 1879)在总结19世纪中叶以前分散的、个别的电磁现象研究成果的基础上,特别是在法拉第工作的基础上,引入位移电流概念,于1864年提出了电磁场基本规律,并用数学方程式表示出来.即是把全部电磁现象归结为两组偏微分方程

$$
\left.\begin{aligned}
\frac{\partial H_z}{\partial y}-\frac{\partial H_y}{\partial z}&=J_x+\frac{\partial D_x}{\partial t},\\
\frac{\partial H_x}{\partial z}-\frac{\partial H_z}{\partial x}&=J_y+\frac{\partial D_y}{\partial t},\\
\frac{\partial H_y}{\partial x}-\frac{\partial H_x}{\partial y}&=J_z+\frac{\partial D_z}{\partial t};
\end{aligned}\right\}
$$

$$
\left.\begin{aligned}
\frac{\partial E_z}{\partial y}-\frac{\partial E_y}{\partial z}&=-\frac{\partial B_x}{\partial t},\\
\frac{\partial E_x}{\partial z}-\frac{\partial E_z}{\partial x}&=-\frac{\partial B_y}{\partial t},\\
\frac{\partial E_y}{\partial x}-\frac{\partial E_x}{\partial y}&=-\frac{\partial B_y}{\partial t}.
\end{aligned}\right\}
$$

式中(H_x,H_y,H_z)是磁场强度矢量在直角坐标系中的分量;(E_x,E_y,E_z)为电场强度矢量在直角坐标系中的分量;(B_x,B_y,B_z)为磁感应强度矢量在直角坐标系中的分量;(D_x,D_y,D_z)为电位移矢量在直角坐标系中的分量;(J_x,J_y,J_z)为传导电流密度矢量在直角坐标系中的分量.这两组方程称为麦克斯韦方程.它们确定了电荷、电流、电场、磁场之间的普遍联系,是电磁学的基本方程.如果再加上初始条件或边界条件就可以确定任何时刻电磁场的状态.1873年,麦克斯韦的《电磁理论》出版,进一步奠定了电磁学理论基础.麦克斯韦的电磁学与牛顿的力学这两大理论已经能够解释物质世界中物体的运动、热、声、光、电、磁等一切物理现象,形成了古典物理学.

对流体运动所遵循的物理规律,在18、19世纪期间有深入的研究.流体根据其物理性质分为粘性与无粘两类.什么是流体的粘性呢?流体虽然不承受切应力,只承受法应力,但对切向变形并不是没有抵抗的,这种抵抗就是内磨擦.流体的内磨擦称为粘性.流体在静止或匀速运动时无相对滑动,这时粘性表现不出来.无粘气体亦称理想气体.对无粘流体运动规律的精确数学描述有欧拉方

程;粘性流体运动规律的精确数学描述有纳维-斯托克斯方程.这两个方程是最基本的,得到了非常广泛的应用.

纳维-斯托克斯方程.它是描述粘性流体动量守衡的运动方程,简称N-S方程.此方程是法国科学家纳维(C. L. M. H. Navier, 1785—1836)于1821年和英国物理学家斯托克斯(G. G. Stokes, 1819—1903)于1845年分别建立的,故名.它在直角坐标系中可写成:

$$\left.\begin{aligned}\rho\frac{\mathrm{d}u}{\mathrm{d}t}&=-\frac{\partial p}{\partial x}+\rho X+\mu\cdot\triangle u,\\ \rho\frac{\mathrm{d}v}{\mathrm{d}t}&=-\frac{\partial p}{\partial y}+\rho Y+\mu\cdot\triangle v,\\ \rho\frac{\mathrm{d}w}{\mathrm{d}t}&=-\frac{\partial p}{\partial z}+\rho Z+\mu\cdot\triangle w.\end{aligned}\right\}$$

式中ρ为流体密度;(u,v,w)为流体速度矢量在直角坐标系中的分量;p为流体各向同性压力;(X,Y,Z)为体积力在直角坐标系中的分量;μ是动力粘性系数;$\triangle$是拉普拉斯算子.这一N-S方程概括了粘性不可压缩流体流动的普遍规律.从理论上讲,N-S方程再加上初始条件或边界条件就可以确定流体的流动状态.当忽略流体粘性(即$\mu=0$)时,则N-S方程就变成通常的欧拉方程

$$\left.\begin{aligned}\rho\frac{\mathrm{d}u}{\mathrm{d}t}&=-\frac{\partial p}{\partial x}+\rho X,\\ \rho\frac{\mathrm{d}v}{\mathrm{d}t}&=-\frac{\partial p}{\partial y}+\rho Y,\\ \rho\frac{\mathrm{d}w}{\mathrm{d}t}&=-\frac{\partial p}{\partial z}+\rho Z,\end{aligned}\right\}$$

即是无粘流体运动方程.欧拉(L. Euler, 1707—1783)是瑞士数学家,理论流体动力学的创始人.

奥地利物理学家波尔茨曼(L. Boltzmann, 1844—1906)于1872年提出对气体分子运动论的基本方程:

$$\frac{\partial f}{\partial t} + v \cdot \frac{\partial f}{\partial x} + \frac{F}{m} \cdot \frac{\partial f}{\partial v} = \int (f f_1' - f f_1) g b \mathrm{d} b \mathrm{d} \varepsilon \mathrm{d} v_1.$$

式中 $f(x,v,t)$ 是质点分子速度分布函数；F 为作用在分子上的外力；m 为分子质量；g 为碰撞前速度分别为 v、v_1 的两个分子的相对速度值；b 为假设第一个分子静止时，第二个分子运动轨迹的渐近线到第一个分子重心的垂直距离；ε 为第二个分子运动轨迹平面同通过第一个分子重心并与相对速度平行的某一固定平面之间的夹角. v、v_1 是碰撞前两分子速度，v'、v_1' 是在那两个分子碰撞后的速度，$f_1 \equiv f(v_1)$，$f' \equiv f(v')$，$f_1' \equiv f(v_1')$. 方程右端称为碰撞积分，代表由于分子相互碰撞引起 f 的变化. 这是一个非线性积分微分方程，用于描述气体分子速度分布函数的变化. 它对研究稀薄气体动力学有重要意义；不过，直接求解波尔茨曼方程十分困难，通常都是对它简化后使用.

上述各个科学基本规律，在经典力学、经典物理学及近、现代的量子力学、流体力学、稀薄气体动力学中具有代表性. 它们的精确数学表达式：牛顿方程、欧拉方程、N-S 方程、麦克斯韦方程、薛定谔方程、波尔茨曼方程，以及由它们派生出的方程，几乎覆盖了一切物理、力学和工程科学领域.

§3 计算机与计算数学推动科学工程技术进步*

计算方法的进步依赖于计算工具的改进，历史上每一种改进的计算工具出现，都曾推动过计算方法的发展. 四十年前世界第一

* 在本节中请参考下列文章：
冯康、周毓麟，第三手段 —— 科学与工程计算，《科技日报》，1987.7.4.
石钟慈、李德元，科学与工程计算及科研与工程设计，《科技日报》，1987.7.4.
曾庆存，大气科学中的数值模拟，《科技日报》，1987.7.6.
崔俊芝、周天孝，有限元方法与结构工程，《科技日报》，1987.7.6.
李松波，气动计算 —— 获得气动力数据的第三手段，《科技日报》，1987.7.6.
黄绪德，石油勘探中的大型计算需求分析，《科技日报》，1987.12.28.

台电子计算机 ENIAC 在美国诞生，它每秒能作三五百次运算，程序控制完全自动，比当时机械式计算器一下子快了一万倍. 这是人类计算工具进化史上的一次具有划时代意义的飞跃. 它对计算工作的推动作用，是历史上任何一次计算工具的改进所不能比拟. 从此，数学中的代数方程、微分方程，特别是那些科学基本规律的精确数学表达式，以及概率统计等的数值求解成为可能，而这些数学问题正是科学与工程上急待解决的. 在生产发展与科技进步的推动下，计算方法研究工作有了突飞猛进的发展. 使用电子计算机自动解算问题的程序设计方法也建立和发展起来了. 于是，计算数学在数学中形成了一个独立分支. 计算数学的内容大致分为计算方法和程序设计两方面，主要研究数学和逻辑问题怎样由电子计算机加以有效解决. 目前由于生产实践与科学实验提出的数学问题愈来愈复杂，以及电子计算机的功能不断在扩大提高，有力地推动了计算数学的发展.

电子计算机与计算数学的发展史虽然仅有四十余年，其应用却已很广泛了，并对科学与工程技术的进步起到了非常大的推动作用. 这可以从以下几个方面得到了解.

1. 有限元方法是计算机与计算数学发展中的一项重大研究成果，是一种效率很高的科学工程计算方法. 50 年代末和 60 年代初，在欧美和中国沿着不同的学术道路，各自独立创造了这一方法. 有限元方法是一门边缘性的交叉学科，多方面的专家学者为它的创立、发展和推广应用作出了贡献，冯康为这一科学计算方法的创立和发展作出了独特的贡献. 这一方法特别适用于物理上、几何上复杂的结构，又易于编制程序，易于在计算机上计算. 三十年来，有限元方法的应用获得了巨大成功，在计算机配合下，普遍用来设计水坝、桥梁、高楼、飞机、汽车、船舶、电机等，有限元程序业已成为各工程分析的日常工具. 有限元方法应用于科学、工程与生产实际，产生的经济效益是很高的. 例如，80 年代初，我国建造四川龚嘴水

电站时，运用有限元方法对混凝土重力坝的弹 性接触问题进行了解算，解决了大坝在蓄水条件下进行纵缝灌浆的技术和控制数据. 这不仅避免了放完水再灌浆而停电半年至一年所造成的经济损失和巨额搬迁费，而且每年枯水期可多发电 1.2 亿度，约合标准煤 5.1 万吨. 有资料指出，目前在美国使用有限元软件的直接经济收入 每年达一亿美元. 在创造新颖设计、加快设计进度、降低工程造价、加快工程进度、提高工程安全度等方面所获得的间接经济效益和社会效益就更可观了.

计算数学发展中的这一重大进步，还可从历史对比看出它的作用和意义. 12 世纪起在西欧兴起了哥特式大教堂的建造. 当时人 们懂得几何学和美学，但不大懂力学，人们采用原始的办法（所谓 试错法）来设计建造，即边设计边施工，结构垮了，再修改重造. 人们经过两个世纪的失败与成功的经验教训才掌握到其中的科学技术. 近代力学和技术科学兴起了，人们发展了比较先进的实验方法，同时又建立了完整的基础理论方程. 但是由于解方程的困难，对 于复杂结构通常还不能提供精确的定量的指导. 计算机和有限元 方法的出现才使情况根本改观. 有人曾用有限元程序考查了哥特式大教堂结构的稳定性，高度从 104 英尺算到 134 英尺只用几分钟时间，把应力和位移分布弄得清清楚楚，从而知道了低结构是安 全的，高结构的顶部要垮，如遇每小时 90 英里风速负荷其低结构 也要垮，使人类在工程设计上得以从必然王国向自由王国迈出了一大步.

2. 促进和发展数值天气预报. 传统的天气预报工作，是对天气进 行人工观测，积累资料，在此基础上再加以分析和归纳，找出它的 变化规律来，进而提出假设和理论，制作天气预报. 根据这种简略 的定性办法所做出的理论解释是比较粗糙的，做出的天气预报常 常是不够准确的. 科学的预报方法是根据科学基本规律的精确数学表达式进行定量计算，并用以 指导和提高理论分析工作，然

后做出天气预报.这里说的科学基本规律的精确数学表达式是指天气预报方程及其边界条件,所谓定量计算是指求解这一数学问题.本世纪20年代初人们即已列出了天气预报方程,它是一组非线性偏微分方程,规模大,很复杂,用数学分析方法尚无法解算,只能采用数值求解的办法.对天气预报方程数值求解也叫做对天气的数值模拟.在20年代里由于没有相应的计算工具,对如此复杂的偏微分方程根本无法进行数值解法研究.直到50年代有了电子计算机,计算数学发展起来了,才开始进行这一数值模拟研究,延迟了三十余年.研究工作的初期是在电子计算机上进行天气预报计算试验,并很快完成了由定性描述到定量计算的突破,再经过一段时间的完善和改进,终于建立了数值天气预报业务.现在世界各国在天气预报中普遍采用了数值预报的方法,从此人类在了解和掌握天气方面才前进了一大步.

数值天气预报业务正在向更长和更短时效两方面延伸,即要做出像一个月时效的长期预报和像一天内的台风和暴雨等灾害性天气的短期预报.

气候预测更需要采用数值模拟方法.气候指的是相当长一定时段(如一个月、一个季度、一年或十年)内天气的平均状态,它不像天气形势预报那样有天气系统可寻,气候变动趋势的规律更加隐蔽,控制因素也更多.近一二十年来,对气候的数值模拟研究很活跃.例如人们共同关心的由于人类活动所造成的二氧化碳和其他污染物质增加所形成的气候效应问题的数值模拟研究就取得了很大进展.预测到今后地球表面的气温将普遍升高,甚至会引起南极的冰山融化淹没一些陆地,危害人类的生活.气候预测直接关系到国计民生,各国政府对此非常重视,现已制定出一个世界气候研究计划,是个国际合作项目,计划在今后一二十年内逐步解决三种时间尺度的气候预测问题.一是月或季节尺度;二是年际尺度;三是年代际(十年以上)尺度.月或季节尺度指的是长期天气预报,

年际尺度指的是气候异常预测，年代际尺度指的是气候趋势预测.我们相信，在不远的将来对一些影响人类生活较大的气候异常事件和十至几十年乃至百年气候变化趋势，是能够做出较为符合实际的预测.

气候和地理环境的关系极为密切，在地理科学中也将会越来越多地建立数学模式和使用数值模拟方法，逐步形成一些定量计算的新分支学科.

数值模拟研究在海洋科学中有很广阔的前景.例如海洋调查，直接调查无论在技术上或设备上的难度都很大，投资也大，调查工作本身又有局限性，故不大可行；而采用数值模拟方法不受时空条件限制，又经济方便，故比较可行.

现代大气科学研究实践表明：数值模拟可以大大缩短研究工作周期，使检验和发展理论工作进行得更快更好，特别是它可以模拟在实验室内无法进行的实验，是发展大气科学和地学的一个重要手段.未来在地学上的突破，很可能是以现有的观测事实为依据，通过数值模拟研究来实现.

3.油气资源的勘探与开采，愈来愈多地倚重于数值模拟的手段.当今世界各国都把新型高性能计算机优先用于石油工业上，应用于这方面的算法研究工作也越来越活跃了.地震勘探方法产生于本世纪 30 年代，到 50 年代由于电子计算机的应用和计算数学的发展，形成了一整套综合勘探方法，这是当前油气资源勘探的主要方法.所谓地震勘探，即人工激发地震波在地下半空间传播，并在地表收取散射反响信号，据此来反推可能含油构造与岩性.这在数学上相当于求解地震波动方程的某种反演问题.勘探数据处理的精确化对探明贮量，提高打井的准确率有重要的指导意义.人们认为勘探事业中对于计算机的投资是花费最少而收益最大，减少一次打错井就回收了一台计算机的费用.

在油田开发中，特别对二次或三次回收开采中，计算机与计算

数学的应用也愈来愈重要. 油田开采的实地试验虽然是可行，但是代价是数值模拟的千百倍，而且耗时经年，结果还不太精确，对比之下，人们更加注重数值模拟手段. 当前在数值模拟上正从几何上简单的二维模型过渡到复杂的三维模型，在物理上也从单一的“黑油”模型过渡到更为完整的多相“组合”模型. 随着计算机的向量化和并行化以及计算方法的改进，这种现场规模的精确计算将临近于实现. 美国曾有人估计如能提高 10% 一 20% 的回收率(指二次或三次采油)，经济效益可达千亿美元，相当于发现一个阿拉斯加北坡或北海油田.

4. 数值模拟对于核电站的设计已经起了很大作用，这种手段今后对于核电站的安全运行将同样起重要的作用. 在 70 年代，美国洛斯阿拉莫斯国立研究所曾初步研制出了核电站事故初期分析的计算程序. 三里岛核电站事故发生后，美国官方委托上述研究所利用该程序作计算模拟对事故进行分析. 计算重现了事故后最初三小时的基本过程. 算得的压力、温度与水位同事故过程中发生的结果吻合非常好. 特别重要的是算出了燃料棒外壳的最高温度，这是实测中无法得到的信息. 据此可以估算出燃料棒外壳的氧化量，并由此释放出的氢气量，搞清了事故的关键. 这是一个用计算手段得到了解决而用实验手段解决不了的问题.

这项计算是在 CDC7600 大型计算机上进行的，花了 20 个小时. 鉴于模拟计算对核反应堆安全运行的重要意义，人们建议将这类模拟计算纳入运行，以便操作人员在事故发生时评估应采取哪个应急措施为好. 为此，计算必须在几分钟内完成，这就需要装备比 CDC7600 运算速度快 1000 倍计算机，或者从改进算法着手，从算法上提高功效 1000 倍.

5. 气动计算促进了飞机、导弹设计工作上的一场变革. 飞行器(航空飞机、人造卫星、导弹、航天飞机的通称)的设计需要靠气动力数据，获得气动力数据，一靠风洞实验，二靠自由飞，三靠气动计

算.气动计算是指在电子计算机上解算气动方程(欧拉方程或N-S方程)和飞行器流场的数值模拟计算.流体运动的科学基本规律的发现及其精确数学描述——欧拉方程与N-S方程早在150年前就问世了,但应用它解决工程中的实际问题却延迟到本世纪50年代或60年代初才开始,是在有了电子计算机以后,计算数学兴起,一些有效的差分格式和稳定法等计算方法相继问世使得欧拉方程乃至N-S方程才开始应用起来.

在航空工业中,风洞的发展已逾80年.它是飞机设计的主要手段.80年来,每当设计一架新型飞机,风洞实验量增加了1000倍,单位时间的实验费也增加了1000倍,经济负担愈来愈重,况且,这种物理手段本身有着很大的局限性.在航空工业中气动计算的发展才20年,在这20年内,如以固定计算方法来核算,计算成本降低了1000倍;再以固定计算机来核算,而计算方法的效率又提高了1000倍,前一种效率的获得是基于计算机的进步,后一种效率的获得则是由于算法的改进,二者相辅相成,使气动计算的潜力无穷.对比之下,气动计算应该是很可取的.但人们的认识需要有一段过程,习惯势力常常主宰着事物的发展.据美国官方透露:近年来若干机种设计上的缺陷,常常是等到试飞阶段才暴露,导致对飞机的降格使用或在修补上付出昂贵的代价,甚至报废,不得不重新设计;而这些问题本来是可用当前已掌握的气动计算手段来避免的.采用气动计算手段设计飞机已有两个成功的先例:美国波音767飞机和欧洲空中客车A310飞机的机翼就是采用气动计算设计的.它们比先前单靠实验手段设计的同类机种减少了飞行阻力,节省燃料20%.以400架飞机飞行15年,每加仑汽油1.3美元计,效益可达100亿美元.设计工作本身也会节省经费和缩短周期,整体效益更是显著了.当前国际上已开始了“整架飞机”的气动力计算,相对于传统的“分片”计算又是一大进步.要使完整的粘性流方程与湍流计算得以实现,计算机还需提高计算能力三四

个数量级，估计到20世纪末以前可以实现. 总之，尽管当前在飞机设计的手段中物理风洞仍占主导，但未来发展的趋势定会倾向于“数值风洞”(即气动计算).

在航天工业中，卫星、导弹与航天飞机的设计工作也呈现出注重“数值风洞”的趋势，因为用“数值风洞”算出来的数据与物理风洞吹出来的数据可以相互印证，使设计工作取得更好的效益.

“数值风洞”与物理风洞相比较，有以下主要优点：

i 通用、灵活、实验周期短，模型与试验条件的修改非常方便.

ii 经济、高效. 计算机价格逐年下降，计算方法效率逐年提高，同一问题的计算经费就大幅度地下降了. 有人估算，如果用60年代中期生产出来的IBM704计算机解算二维时间平均N-S方程需解算30年，花费1000万美元，现在用克雷-1计算机并采用新算法，则只需半小时，费用不到1000美元，效率提高50万倍，成本降低10000倍.

iii 研究成果有良好的可继承性和可扩展性，因为算法和软件在系列化的计算机发展进程中可以永恒使用.

我们相信：由于数值模拟手段促成工程设计中的这场变革，将会有力地推动各国航空航天工业的发展.

6. 核武器的研制必须依靠高效能的科学工程计算手段. 前面说了，飞机的设计尚可采用风洞实验来模拟飞行试验，以获取气动力数据. 对核武器来说，现在还没有什么设备可对核武器产品的性能事先作出试验，对关键性的设计参数也无法在事先通过试验来获取，只有通过核爆炸试验，用测试仪器测得数据. 这种用事后的效应来验证产品是否符合设计思想的做法，耗资大，效果又差. 因为核爆炸过程与核反应过程都是在高温高压下进行的，温度高达几千万摄氏度，压力高达几百万大气压，核爆炸的巨大能量又是在微秒量级的时间内释放出来的，在一次核试验的瞬间要求测量出

核武器内部的细致反应过程是十分困难的. 况且，进行一次核试验，不仅从技术上、经济上、人力上要付出很高的代价，就是从国家的政治路线方针来考虑，也只能有限制地进行. 诚然，数值模拟已经成为研制核武器的必要手段. 核反应是一复杂的物理过程，描述它的数学模型是一组非线性偏微分方程，再加上定解条件. 这组方程的解能给出核爆的各个细节的图象，以及定量的数据和运动的全过程. 当然，对如此复杂的非线性偏微分方程组只有求助于数值方法求解. 即采用数值模拟手段来了解核爆炸时运动的全过程，以及获取设计上所需的数据. 进一步说：核爆炸试验只能提供一些综合效应的数据；数值模拟却可以看出各种因素和机制是如何相互影响而起作用的，从而可以了解到核反应的运动规律并掌握产品设计的规律，这对于发展核武器是至关重要的. 当然用数值模拟完全代替核试验是不可能的，但可以减少试验次数.

核反应的物理过程本质上是一个三维问题，需要建立三维数学模型进行模拟计算. 当前限于计算机的条件，只能采用简化的一维或二维数学模型予以近似模拟. 改进物理、数学上的理论方法及提高计算方法的精确度，是提高核武器设计技术的两个最基本的手段. 仅就改进算法来看，改进有限元网格法可以非常有效地提高模拟计算的精度；如能采用并行计算方法将会使计算速度提高近千倍. 当然，改进计算机的性能（扩大容量与加快运算速度）也是提高计算精度满足设计需要的一个途径. 利用计算机提高精度可从两个方面下手：一是采用三维的数学模型；二是加密网格点. 实际的经验表明：当取 100 个网格点时，二维计算所需的计算机容量是一维的 200 倍，运算时间也是 200 倍；三维计算所需的计算机容量则是一维的 30000 倍，运算时间也是 30000 倍；当取 1000 个网格点时，二维计算所需的计算机容量是一维的 2000 倍，运算时间是 1000 倍；三维计算所需的计算机容量是一维的 300 万倍，运算时间是 10 万倍.

7. 计算机与计算数学在综合利用、通盘规划与提高全局经济效益方面起着重大作用. 物资调拨、运输调度、厂矿布局，以及系统工程方面每一个方案的改进都将导致千万元甚至上亿元的效益. 这方面提出的数学问题往往是大规模的(非)线性最优化问题，它的求解只是在有了高性能的计算机和高效率计算方法以后才具有实际可能性. 例如，电站的最佳潮流问题，它具有重大的经济意义. 问题提出已逾 20 年，然而 60 年代沿用的算法又慢又不可靠，不可能用来解决这样的大型问题. 1984 年系列二次规划方法应用成功，计算效率上有突破性提高，这才解决了先前认为难以处理的大规模最佳潮流问题，引起了电力界极大的关注. 1985 年加拿大多伦多的安大略水电站(350 汇流条)经过实时分析，对应用最佳潮流程序进行电压控制给出了评价：有功损耗减少 3.7%，每年合 250 万美元，无功损耗减少 19.7%. 总之，由于计算机和算法的改进，程序运行的速度很快，从而使电站最优电压控制的实时应用得以实现.

8. 计算机的不断完善与计算数学的发展，对科学实践与生产实践带来了革命性的变化，这不仅是因为它能够为科学实践与生产实践提供计算工具，从而产生了巨大经济效益；而且是因为它为理论模型与生产模型的创新提供了一个高效的数值模拟或数值实验的手段，还为实验研究提供了一个多点、在线、实时数据采集与处理系统，使人们得以深入地了解与认识事物变化的真实过程. 从前面列举的诸例可以看出：计算的方法正越来越广泛地应用在科学研究与工程设计的各个领域，在当代的技术发展中起决定性作用，并且直接或间接地影响到一切工业、经济和基础科学部门. 我国与美苏等国家的建设经验表明：科学工程计算是关系到国家安全、经济发展与科技进步的关键性因素.

今天业已存在的现实的可能性是把自牛顿起的 400 年间人类所积累的精确科学技术的知识和智慧转化为计算机程序用以去勘

探与开发自然资源，设计制造出更先进的产品，也就是通过计算的手段把这一科学宝库转化为生产力，把精神财富变成物质财富.

§4 计算机把计算的方法推向人类科学活动的前沿

近代科学方法论的起点是观测，哥白尼日心说的创立，维萨里血液循环“以心脏为中心”的发现，以及后来的开普勒行星运动三定律的发现，等等，都是通过观测和实验来认识物体的本来面貌和它们的基本运动规律. 但科学方法论真正受到人们的重视是从伽利略开始的，是伽利略在人类科学活动中系统地引进了科学实验方法. 古代人是从哲学思想产生假说，并用来解释现象；而近代科学则是在观测与实验的基础上归纳出定律、法则，能够说明过去不能说明的种种现象，能够揭露和纠正以前的种种愚昧和谬误，使人类社会得以从漫长的黑暗与愚昧时代中走了出来. 这是一大进步. 如果说亚里士多德是从抽象的总体上进行推理来认识世界，那么伽利略则是把自然现象分解为单一因素，通过实验[illegible]立因果关系，而后用数学加以描述作为假说，再进一步通过实验来验证，使之逐步逼近自然界自身情况，达到构成特有的规律. 这是科学方法论的重大变革.

牛顿的贡献则是在前人哥白尼、伽利略等对自然现象分解的、孤立的一系列重大发现的基础上，把物体运动（包括行星的和地面的）作了整体统一的考察，并采用数学方法来描述物体运动过程中的因果关系，从而建立起具有完整体系的力学物理学，奠定了力学物理学的理论基础，并在人类科学活动中树立起了理论方法的典范.

自伽利略系统地引进科学实验方法和牛顿奠定力学物理学的理论基础以来，近代科学方法论就分为实验与理论的两大环节. 这

是人类在科学活动中两个行之有效的手段，沿用至今.

20 世纪 40 年代发明了电子计算机，这一计算工具的发明，其意义不同于历史上的算盘与计算尺的发明，也不同于望远镜和显微镜等其他科学工具的发明. 电子计算机延伸和强化了人的思维智能，它可以影响人类所涉足的一切科学技术领域. 电子计算机刚诞生时，知名美国数学家冯·诺伊曼(John von Neumann，1903－1957) 就深刻指出这一新工具拥有巨大潜力和计算的方法作为一种新的第三种科学方法的发展前景. 其后，计算机的飞速发展正把计算的方法推向人类科学活动的前沿，使它上升为一种主要的科学方法，特别对于科学的定量化起了重要的作用. 事实表明：实验、理论、计算已成为科学方法上相辅相成而又相对独立的三个环节. 如果说过去是沿用理论⇄实验的方法，如今人们习惯采用理论⇄计算⇄实验的方法了. 也就是说，人们从已有的理论出发，将物理现象、物理规律描述为一系列数学方程式，提出数学模型，再在计算机上求出数值解，再以实验的方法予以核对验证，从而进一步探索物理规律，形成设计思想，研制出新产品. 由计算机推动的当代计算方法的兴起，并形成与实验方法、理论方法并立之势，是伽利略、牛顿以来在科学方法论上的最大进展，推动着科研实践中一场深刻的，不可逆转的变革.

§5. 偏微分方程数值解法及在科学与工程中的应用

"偏微分方程数值解法及在科学与工程中的应用"是"现代数学中若干基本问题的研究"项目中的一个课题. 科学与工程技术中许多重要问题的研究最终归结为一组微分方程的求解. 而数学分析的方法对于微分方程的实际解题能力非常有限，只限于小规模、线性、均匀性与几何规则性的"初等"情况. 然而，科学与工程

技术中提出的微分方程大都是大规模的，非线性的，非均匀性的和几何非规则性的. 数学分析的方法对它们则完全无能为力，使用现代计算机进行数值求解是解决这一类问题的唯一途径. 因此，偏微分方程数值解法的研究，是关系到国民经济建设和国防建设各个重要部门的计算机应用方面的最重要课题之一. 世界各发达国家都在这方面投入了大量的人力和财力. 历来超级计算机的研制都以解决这方面的问题为主要背景. 所谓超级计算机，是指特定时段内接近当时技术极限的计算机系统，其功能超过当时市场上销售的商品计算机. 偏微分方程数值解法及在科学与工程中的应用是国际上非常活跃的前沿研究课题.

这一课题的目的在于组织国内在这方面力量比较强的五个单位：中国科学院计算中心、北京应用物理与计算数学研究所、北京大学数学系、清华大学应用数学系、武汉大学数学系，并由冯康担任课题负责人，就以下七个问题

Ⅰ 哈密尔顿动力学的辛几何算法；

Ⅱ 有限元方法的理论及其应用；

Ⅲ 解具奇性问题的高精度算法；

Ⅳ 近代物理学中的偏微分方程的差分方法的研究；

Ⅴ 非线性偏微分方程和随机微分方程数值解；

Ⅵ 奇异摄动问题的数值方法；

Ⅶ 新一代计算机解偏微分方程的计算方法

进行研究，想从基础理论与基础应用方面解决国家在经济、国防与科学中的计算问题.

当今，偏微分方程数值求解方面的前沿研究课题，大致可分为两类：一类是寻找更加有效的算法来解决“老”问题；另一类是对于工程物理中的“新”问题设计新的算法. 本课题则是在上述七个问题上，就这两类研究开展工作. 下面分别予以介绍.

1. 哈密尔顿*动力学的辛几何算法**　自然界几乎所有现实的无耗散的动态过程，包括无粘流体力学动态问题，本质上都可纳入哈密尔顿形式体系，数学上由哈密尔顿方程来描述．冯康曾提出如能充分认识哈密尔顿系统的深刻数学内涵，从此出发建立其正确的计算方法会导致动态问题的合理解决．并于 1984 年在第五届国际微分方程与微分几何学术讨论会上以及在他和他的研究组的工作中创导了哈密尔顿动力系统计算方法的系统研究和基于辛几何理论的研究途径，形成了辛几何差分格式的系统构造方法，建立了相应的基础理论框架．这既为计算数学也为哈氏动力系统理论开创了新方向新领域，得到国际学术界包括许多知名科学家的确认和高度评价．这一新方向处在计算数学、动力系统，以及非线性力学物理学领域的交叉点．可以预计在不远的将来将有重大的进展和广阔的应用前景．要研究的具体问题有：

Ⅰ　研究有限维数和无限维数的哈氏动力系统的数学性质；

Ⅱ　研究哈氏动力系统的辛几何计算方法及其基础理论；

Ⅲ　辛几何计算方法对于各类有限维或无限维的哈氏动力系统的应用拓广．

2. 有限元方法的理论及其应用　有限元解的误差估计无论是理论上和应用上都极为重要．以往仅讨论先验估计，但这不足以解决实际计算结果的误差分析问题．我国学者在 70 年代后期提出通过逐步加密网格法得到逐点后验误差，外国在同一时期也发展了一种使用能量基数的后验误差估计．

---

* 哈密尔顿（W. R. Hamilton，1805 － 1865），英国数学家、物理学家．1834 年建立“哈密尔顿原理”，使各种动力学定律都可以从一个变分式推出，并使人们看到了力学和几何光学的相似之处，后来推广到物理学的许多领域，如电磁学等．他把广义坐标和广义动量都作为独立变量来建立动力学方程，这种形式的方程称为“哈密尔顿正则方程”，也就是哈密尔顿系统．

** 辛几何：symplectic geometry，耦对几何．

有限元解的渐近展开是误差分析的一个十分重要有用的工具，由此可得出一些方法的超收敛性并可导出许多高精度的有限元格式.我国学者在此方面首创性的工作很有特色，在国际上已引起广泛重视.

非标准有限元是有限元法应用中的一项重要进展.它通常从力学直观出发建立有限元模式，往往能以较少的计算量而获得较高的精度，但理论分析十分缺乏和困难.我国学者在此方向上获得一系列重要的成果，为一些很有用的非标准有限元建立了可靠的理论基础.

边界元法是有限元发展中的一个重要阶段.对某些问题它比有限元更为有限，将有限元中的自适应技术应用于边界元，这是一项很新的研究课题.我国学者在正则归化边界元法方面取得了一系列成果，再加上自适应技术，二者结合，很有新意.具体将解决以下六个问题：

Ⅰ　有限元解的渐近展开及超收敛性研究；

Ⅱ　有限元解的后验误差估计(各种基数意义下)；

Ⅲ　自适应加密方法和外推技术；

Ⅳ　非标准有限元研究；

Ⅴ　N-S 方程的有限元解法；

Ⅵ　边界元方法及自适应技术.

3.解具奇性问题的高精度算法　许多非线性双曲型方程组、抛物型方程组、抛物-双曲耦合组、双曲-椭圆耦合组的解，通常带有某种奇异性.求解这类方程的已有一些算法，其精度不高，不能满足需要.因此，寻找高精度算法仍然是国际上非常活跃的前沿研究课题之一.从前，我国学者在研究超音速绕流的数值计算过程中，发展了一个称为“分离奇性”的差分方法.此方法的优点在于花费不多的机时，便能计算出高精度的结果.前几年又将此方法推广应用于若干其他类型的两个自变量问题上，取得很大成功.这次

准备以航天航空工业中提出的跨音速流动问题、开采石油中的二次采油及相变问题以及燃烧问题和生物数学中的某些问题为背景开展三个自变量的偏微分方程组的分离奇性法的研究，以期构造高精度的算法，使得在现有的计算机装备条件下能够解决从前无法解决的问题，满足科学与工程上的需要.

4. 近代物理学中偏微分方程的差分方法研究 广泛出现于近代物理学中的新型偏微分方程，包括弹塑性流体力学方程、Schrödinger 方程、铁磁链方程、Kdv 方程，以及 Zaxapoв 方程等，其中不少是属于具有不同形式和不同程度退化性质的抛物型方程式或方程组. 这些方程是描述近代物理学中的高速射流和侵彻问题、激光 - 等离子体相互作用问题、核聚变问题、混沌问题以及强湍流等问题. 而这些物理问题均是国民经济和国防建设中急需解决的问题. 由于上述问题极其复杂，要对这些问题在理论上作出比较完整的探讨，至少在现今是十分困难的；而数值模拟则能对其内在的变化过程进行完整而又细致的描述，使人们得以认识这些问题的内在机理和规律. 数值模拟可以对许多实际问题进行优化设计从而带来显著的经济效益. 上述方程的非线性和退化性质，要求计算方法得有新的设计和处理，这样做对完善差分方法的理论研究亦有其理论意义. 高速射流和侵彻问题等的研究已受到国内外学术界的极大重视，国外的研究和应用已领先于我们，但多数问题尚未涉足，或涉足不深. 这是一块尚新的研究领域，又关系到经济国防的发展，我国必须组织力量、提供条件开展这方面的研究工作，在近期内取得成效. 在本课题中要解决的具体问题有：

Ⅰ 弹塑性流体力学差分方法；

Ⅱ Schrödinger 方程、铁磁链方程、渗流方程、Kdv 方程、Zaxapoв 方程的差分方法及其理论论证；

Ⅲ 超热电子型的退化抛物型方程以及在区域边界上退化的抛物型、双曲型方程的差分方法及其理论论证；

Ⅳ　计算格式的并行算法；

Ⅴ　计算方法和程序的可靠性及精度的检验与分析.

5. 非线性偏微分方程和随机微分方程数值解　自然科学中的各种基本规律不少是以非线性偏微分方程为其表现形式，特别对一些重要物理现象的研究，如湍流、边界层、燃烧、激波、流动不稳定性等更离不开对非线性偏微分方程进行数值计算和理论分析. 这些问题的研究不仅有着理论意义，而且具有重要的现实意义.

不同学科间相互渗透和结合是现代计算数学发展的一个重要趋势，其中概率统计方法在解微分方程中的应用就是一个典型例子. 随机微分方程在高雷诺数流动、奇摄动、流体计算中的成功应用，日益引起人们的关注. 我国学者近年来在偏微分方程近代数值方法的理论和应用上取得了可喜成果，其中包括流体力学的有限元与无限元方法，N-S 方程有界区域的分步法，高雷诺数流动的涡团法，燃烧流动的间断解理论和数值计算，统计力学中随机微分方程的数值模拟等. 当前要研究的主要问题有以下六个：

Ⅰ　化学反应流动的计算方法和数值模拟；

Ⅱ　拟线性双曲型差分方法；

Ⅲ　有限元和无限元方法；

Ⅳ　欧拉方程的涡团法；

Ⅴ　随机微分方程方面的随机激动法在高雷诺数流动的应用，以及统计物理和力学中随机激动模型的理论和数值模拟；

Ⅵ　N-S 方程分步法.

6. 奇异摄动问题的数值方法　在力学、物理和工程科学中经常遇到含有小参数的问题，其中高阶微商项带小参数的微分方程定解问题尤其重要，称奇异摄动问题，例如高雷诺数不可压粘性流体的流动，对流占优的热传导问题，半导体器件方程组等. 奇异摄动问题的分析理论经过力学家、数学家及其他学科领域学者几十

年的努力，目前已成为应用数学的一个重要分支. 近十年来有不少学者从计算数学的角度研究奇异摄动问题，通过大量数值实验发现，当小参数很小时，即使是对最简单的奇异摄动问题用所有成熟的计算方法也得不到比较准确的数值解. 因此，必须发展适合于奇异摄动问题的计算方法和格式.

主要研究内容是：

Ⅰ　研究偏微分方程奇异摄动问题的数值格式和有效的计算方法以及理论分析；

Ⅱ　自适应技术应用于奇异摄动问题；

Ⅲ　对流占优的热传导问题等若干实际问题的研究.

7. 新一代计算机解偏微分方程的计算方法　数学物理中的许多问题可归结为高维非线性偏微分方程，求解这类方程的难度和规模都越来越大. 这就不仅需要寻找可行的计算方法，而且还希望所提供的算法适用于体系结构并行化的新一代计算机. 由于CRAY 等新一代巨型计算机的出现，数值解偏微分方程的并行算法研究日趋活跃. 欧、美、苏等国都有一批科学家在开展并行算法研究. 我国学者基于区域分裂原理创造了一类新型的解偏微分方程的异步并行算法，结合应用取得了不少很好成果，在国内外交流中博得好评. 当前准备开展以下七个问题的研究：

Ⅰ　对高维非线性偏微分方程，进一步研究区域分裂法，包括直接解法与迭代解法；

Ⅱ　提出并行 Schwarz-MG 方法的算法与理论，进行自适应并行化数学软件的基础研究；

Ⅲ　开展并行化有限粒子法研究；

Ⅳ　研究自由表面的计算与自由表面的最优设计，拟解决计算稳定性问题和最优设计的可能性问题；

Ⅴ　变分不等式的投影算法；

Ⅵ　变分不等式的对偶算法及降维算法；

Ⅶ　偏微分方程半离散化后的Stiff方程的算法研究与理论分析.

在本文的后面附有微分方程数值解法及其在科学与工程中的应用方面的有关专著和文献，以便读者深入了解这一领域的发展动向.

§6　结束语

计算，作为第三手段使科学方法论益加完善．这意味着，今后，人类认识自然和改造自然的活动必将以更快的步伐向纵深发展.

在当今国际激烈竞争的形势下，在技术革命浪潮的推动下，一个国家对科学计算的掌握程度是它的自我科技实力的真实度量．科学工程计算既是一门当前可以见效的应用性学科，又是一门有长远战略意义的基础性学科．它与我国的四个现代化建设有广泛、直接、紧密的联系，因此特别值得受到各方面的重视和大力支持.

生产与科学技术的发展，特别是在计算机的不断更新换代浪潮的推动下，科学与工程计算这一领域将越来越宽阔，它会吸引众多有识青年献身于现代科学工程计算研究.

参　考　文　献

1. Feng Kang, On Difference schemes and symplectic geometry, Proceedings of the 1 9 8 4 Beijing Symposium on Differential Geometry and Differential Equations—COMPUTATION OF PARTIAL DIFFERENTIAL EQUATIONS, Ed. Feng Kang, Science Press, Beijing, 1985, 42—58.
2. Feng Kang, Difference schemes for Hamiltonian formalism and symplectic geometry, JCM, 4:3(1986), 279—289.
3. Feng Kang , Symplectic geometry and numerical methods in fluid

dynamics, Lecture Notes in Physics, V. 264, Proceedings of the 10'th International Conference on Numerical Methods in Fluid Dynamics, Beijing, 1986, ed. F. G. Zhuang and Y. L. Zhu, Springer-Verlag.

4. Feng Kang and Qin Meng - zhao , The symplectic methods for the computation of Hamiltonian equations , Proc . of lst Chinese Conf . on Numerical Methods of PDE's, March 1986, Shanghai, Lecture Notes in Mathematics, No. 1297, 1—37, ed. Zhu You-lan and Gu Ben-yu, Springer, Berlin, 1987.

5. Feng Kang, Wang Dao-liu, Symplectic difference schemes for Hamiltonian systems in general symplectic structures, JCM 9:1 (1991) 86—90.

6. Feng Kang, Wu Hua-mo, Qin Meng-zhao, Symplectic difference schemes for the linear Hamiltonian canonical systems, JCM 8:4 (1990) 371—380.

7. Feng Kang, Wu Hua-mo, Qin Meng-zhao and Wang Dao-liu, Construction of canonical difference schemes for Hamiltonian formalism via generating functions, JCM, 7:1(1989).

8. Ge Zhong and Feng Kang, On the approximation of linear H-systems, JCM, 6:1(1988), 88—97.

9. Li Chun-wang and Qin Meng-zhao, A symplectic difference scheme for the infinite dimensional Hamiltonian system, JCM, 6:2(1988).

10. Qin Meng-zhao, A symplectic difference scheme for the Hamiltonian equation, JCM, 5(1987), 203—209.

11. P. Ciarlet, The Finite Element Method for Elliptic Problems, North-Holland, 1978.

12. O. Zienkiewicz, The Finite Element Method, Mcgraw-Hill, 1977.

13. 朱幼兰,钟锡昌,陈炳木,张作民,初边值问题差分方法及绕流,科学出版社,北京,1980.

14. Harten , A . , High resolution schemes for hyperbolic conservation laws, J. Comp. Phys. Vol. 49, pp357—393, 1983.

15. P. Woodward and P. Colella, The numerical simulation of two dimen-

sional fluid flow with strong shock, J. Comp. Phys. Vol. 54, pp151—173,1984.

16. Wu Xiong-hua, Huang Dun, Zhu You-lan, Numerical computation of the flow with a shock wave passing through a strong explostion center, Journal of Computational Math., Vol. 1, No. 3 pp247—258, 1983.

17. Harten, A., B. Engquist, S. Osher and S. R. Chakravarthy, Uniformly high order accurate non-oscillatory schemes Ⅱ, ICASE Report No. 86—22 (April, 1986).

18. P. L. Roe, Approximate Riemann Solvers, Parameter Vectors and Difference Schemes, J. Comp. Phys., Vol. 43, pp357—372, 1981.

19. A. J. Chorin, Random choice solution of hyperbolic systems, J. Comp. Phys., 22, pp517—533, 1976.

20. Wu Xionghua, Wang Yong, Teng Zhenhuan and Zhu Youlan, Numerical computation of flow field with deflagration and detonation, J. Comp. Math., Vol. 2, No. 3, pp247—256, 1984.

21. Xionghua Wu and Youlan Zhu, A scheme of the singularity separating method for the nonconvex problem, Computers & Fluids, Vol. 13, No. 4, pp473—484, 1985.

22. Colella, P. and P. R. Woodward, The piecewise-parabolic method (P. P. M.) for gas dynamical simulations, J. Comp. Phys., V. 54, pp174—201, 1984.

23. Яненко Н. Н., Метод дробных шагов решения многомерных задач Математической физики, Наука Сиб. Отд., Новосибирск, (1967).

24. Тихонов А. Н., Самарский А. А., Об однородных разностных схемах, ЖВМИМФ, Том 1, NO. 1, 5—63. (1961).

25. Fu Hongyuan: Explicit Difference Schemes for Generalized Nonlinear Schrodinger Systemes, 《Proceedings of the 1984 Beijing Symposium on Differential Geometry and Differential Equations》, ed. by Feng Kang, pp 153—158, Science Press, China, (1985).

26. Shen Longjun: Convergence Conditions of the Explicit and Weak

Implicit Finite Difference Schemes for Parabolic Systems, Proceedings of the numerical methods for partial differential equations, Shanghai, Springer-Verlag, p. 127,(1987).

27. Li Deyuan, A Difference Method for Degenerate Hyperbolic Equations of Second Order , in 《 Numerical Methods for Partial Differential Equations》, Eds. Zhu Youlan and Guo Benyu, pp 76—97, Springer-Verlag, Berlin,(1987).

28. Li Deyuan, Difference Schemes of Degenerate Parabolic Equations, J. Comput. Math. vol 1, pp 211—222,(1983).

29. Amsden, A. A., The particle-in-Cell Method for calculation of the dynamics of compressible fluids, Report LA-3466, Los Alamas Sci. Lab. (1966).

30. Wilkins, M. L., Calculation of elastic-plastic flow, Report UCRL-7322, Rev. 1. Lawrence Livermore Laboratory. (1969),(1983).

31. Hirt , C. W., Amsden , A. A. and Cook , J. L. : An arbitrary Lagrangian-Eulerian Computing method for all flow speeds, J. Comp. Phys. ,vol. 14, p. 227. (1974).

32. 徐国荣,于志鲁,廖振民,袁仙春,周淑荣,多物质可压缩流体力学的欧拉数值方法,数值计算与计算机应用,vol. 1,p. 163. (1980).

33. 周毓麟,李德元,龚静芳,一阶拟线性方程物理解的计算,数值计算与计算机应用,vol. 1,p. 16. (1980).

34. 李德元,徐国荣,水鸿寿,何高玉,陈光南,袁国兴,二维非定常流体力学数值方法,科学出版社. (1987).

35. Ying Lung-an and Teng Zhen-huan, A hyperbolic model of combustion, North-Holland Mathematics Studies, Vol. 81,409—434. 1983.

36. Ying Lung-an and Teng Zhen-huan, Riemann problem for a reacting and convection hyperbolic system,J. Appro. Theory Appl. vol. 1,95—1[illegible],1984.

37. 滕振寰,应隆安,双曲型燃烧模型方程组的 Riemann 问题,数学年刊,6A,pp. 13—22,1986.

38. 滕振寰,应隆安,Glimm 方法对于燃烧模型组的收敛性,数学年刊,7A,

pp. 315—324, 1986.

39. Ying Lung-an, The viscosity splitting method in bounded domains, Advances Math. vol. 16, 1, 108—109. 1987.

40. Ying Lung-an, Finite element approximation to axial symmetric stokes flow, J. Compu. Math., vol. 4, 1, 38—49, 1986.

41. Ying Lung-an, Infinite element approximation to axial symmetric stokes flow, J. Comput. Math., vol. 4, 2, 111—120, 1986.

42. Teng Zhen-huan, Chorin A. J. and Lin Tai-ping, Riemann problem for reacting gas, with application to transition, SIAM J. Appl. Math., vol. 42, 964—981, 1982.

43. Teng Zhen-huan, Elliptic-vortex method for incompressible flow at high reynolds number, J. Comput. Physics, vol. 46, 54—68, 1982.

44. Teng Zhen-huan, Variable-elliptic-vortex method for incompressible flow simulation, Lecture Notes in Physics, vol. 264, 600—605, 1986.

45. A. M. Илъин, Разностная Схема Для Диффенциального Уравнения С Малым Параметром При Старшей Производной, Матем Заметки. Т. 6, с. 237—248, 1969.

46. R. B. Kellogg, A. Tsan, Analysis of some Difference Approximation for a Singular Perturbation Problem Without Turning Points, Math. Comp., vol. 32, pp. 1025—1039. 1978.

47. E. P. Doolan, J. J. H. Miller and W. H. A. Schilders, Uniform Numerical Methods for Problems With Initial and Boundary Layers, Boole Press, Dublin Ireland, 1980.

48. A. E. Berger, H. Han(韩厚德)and R. B. Kellogg, On the Behaviour of the Exact Solution and the Error in a Numerical Solution of a Turning Point Problem, Computational and Asymptotic Methods for Boundary and Interior Layers, J. J. H. Miller, ED., Doole Press, Dublin, pp. 13—27, 1982.

49. ——, A Priori Estimates and Analysis of a Numerical Method for a Turning Point Problem, Math. Comp., vol. 42, pp. 465—491, 1984.

50. P. A. Farrell, Sufficient Conditions for the Uniform Convergence of a

Difference Scheme for a Singularly Perturbed Turning Point Problem, SIAM J. Numer. Anal. ,vol. 25,pp. 618—643,1988.

51. K. V. Emelyanov, Difference Schemes for Singularly Perturbed Boundary Value Problems, Ball IV,pp. 51—60,1986.

52. C. Johnson, A. H. Schatz and L. B. Wahlbin, Crosswind Smear and Pointwise Errors in Streamline Diffusion Finite Element Methods, Math. Comp. ,vol. 49, pp. 25—38,1987.

53. M. M. Gupta, A Survey of Some Second-order Difference Schemes for the Steady-State Convection-Diffusion Equation, INT. J. Numer. Methods Fluids, vol. 3,pp. 319—331,1983.

54. 康立山,孙乐林,陈毓屏,解数学物理问题的异步并行算法,科学出版社,1985年.

55. 康立山,陈毓屏,孙乐林等,异步并行算法及其应用,分布式并行处理系统探索(第三部分),武汉大学出版社,60—122,1984年.

56. Kang Li-shan (ed.),Parallel Algorithms and Domain Decomposition, Wuhan University Press, 1987.

57. Kang Li-shan, Chen Yu-ping, Sun Le-lin and Quan Hui-yun, The Asynchronous Parallel Algorithms S-COR for Solving P. D. E'. s on Multiprocessors, International J. Computer Math. , vol. 18, No. 2, 163—172,1985.

58. Kang Li-shan, Chen Yu-ping, Sun Le-lin, Quan Hui-yun, A Class of New Asynchronous Parallel Algorithms for Solving Partial Differential Equations, Proceedings of the 1984 Beijing Symposium on Differential Geometry and Differential Equations —— Computation of Partial Differential Equations, ed. Feng Kang, 237—244,1985.

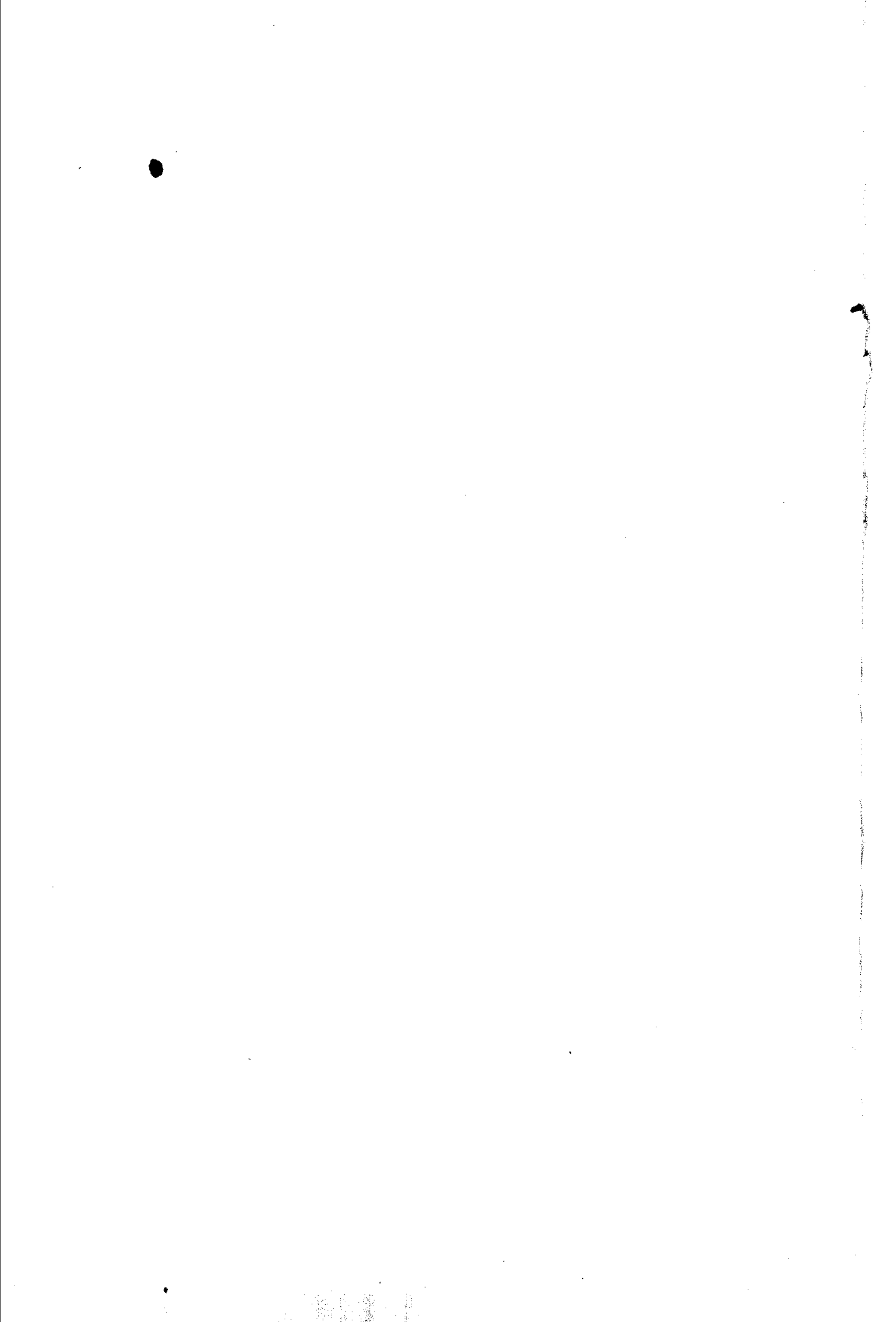
59. H. Troger, 分叉理论——对数学家和工程师的挑战,力学进展,16(4), pp. 535—548,1986.

60. M. Kubicek, M. Marek, Computational method in bifurcation theory and dissipative structures, Springer-Verlag, New york,1983.

61. W. C. Rheinboldt, Numerical analysis of parametrized nonlinear equations John Wiley & Sons, Inc. 1986.

62. H. B. Keller, Numerical methods in bifurcation problems, Springer-Verlag, Berlin, 1987.

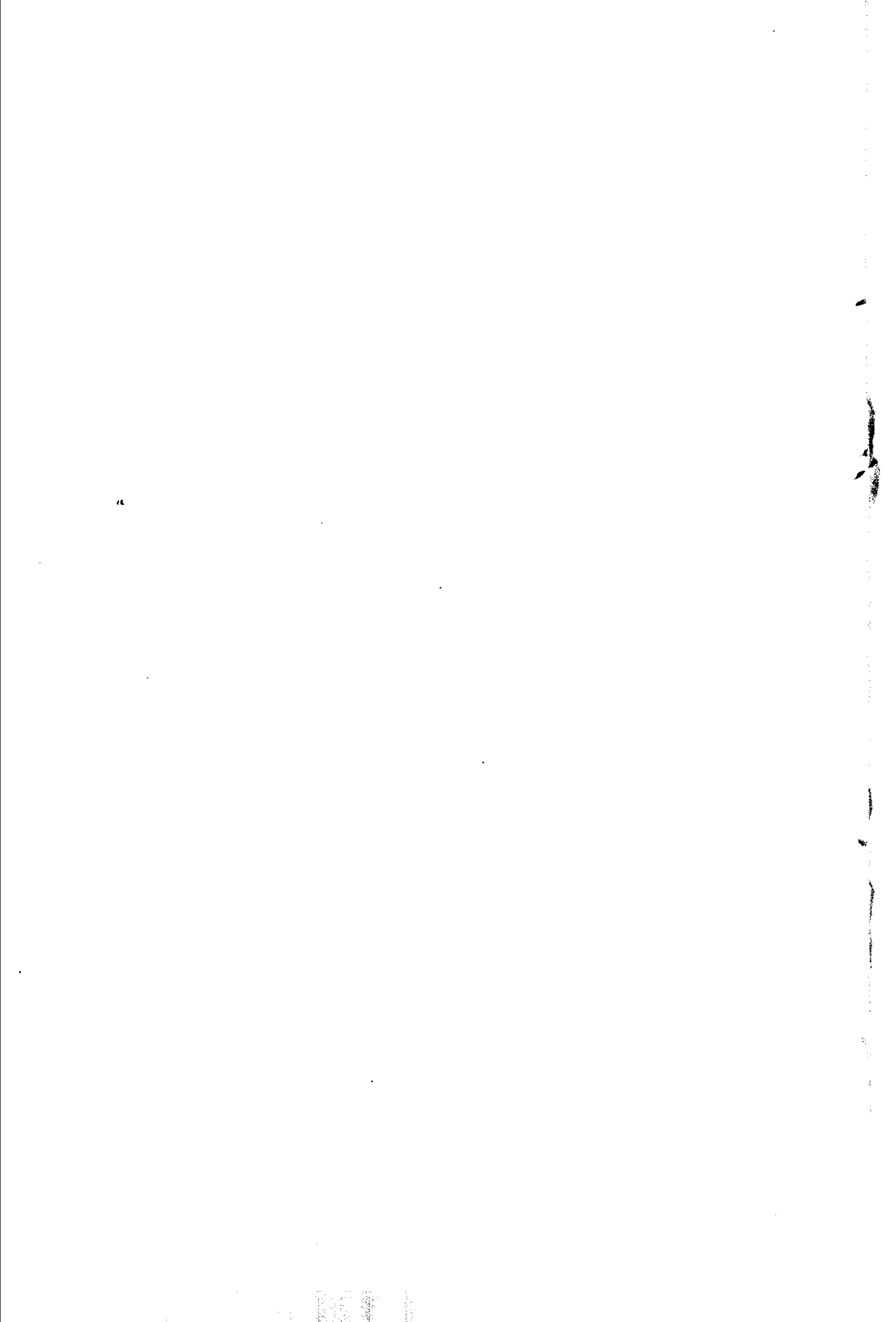
63. H. D. Mittelmann, H. Weber, Eds Bifurcation problems and their numerical solution, ISNM vol. 54, Birkhanser-Verlag, Basel, 1980.



V

整体微分几何及其物理应用

Global Differential Geometry and It's Applications to Physics



谷超豪
（复旦大学数学研究所）
Gu Chaohao
（Institute of Mathematics，
Fudan University）
胡和生
（复旦大学数学研究所）
Hu Hesheng
（Institute of Mathematics，
Fudan University）

整体微分几何及其在理论物理中的应用
Global Differential Geometry and It's Applications to Theoretical Physics

中国数学发展的若干主攻方向
Some Main Aspects on Development of Mathematics in China

整体微分几何是几何学，分析学和拓扑学相互交叉渗透的产物，它的概念和方法，不仅对整个数学的发展起着重要的作用，而且也成为理论物理学的重要工具. 在本文中，我们将根据整体微分几何发展的历史和现状，对上面所说的观点，作一简要的阐述，希望能使大家了解到，为什么整体微分几何学是如此繁荣，如此引人入胜.

§1 微分几何的初始阶段

古典的局部微分几何是研究三维欧氏空间 E_3 的曲线和曲面在一点邻近的性质，它的发展与分析学有着不可分割的联系. 微分几何起源于 17 世纪发现微积分之时，函数与函数的导数的概念实质上等同于曲线与曲线的切线或斜率. 函数的积分在几何上则可解释为一曲线下的面积. 当时，平面曲线、空间曲线及曲面的几何也可作为微积分的应用来了解.

在这方面第一个作出贡献的是瑞士数学家 L. Euler. 1736 年他首先引进了平面曲线的内在坐标这一概念，即以曲线弧长这一几何量作为曲线上点的坐标，从而为曲线的进一步研究奠定了基础. Euler 将曲率描述为曲线的切线方向和一固定方向的交角的变化率. 因而直观上曲率表示曲线的弯曲程度，描述曲线方向改变的快慢程度（相对于弧长而言）. 在曲面论方面，Euler 也有重要的贡献，例如引进了曲面上的法曲率、总曲率、有名的 Euler 公式及球面映照等. 如所知，测地线是平面上直线在曲面上的推广，Euler 和 J. Bernoulli、D. Bernoulli 一起最早地把测地线描述为某些微分方程的解. 在物理的推动下，1736 年，Euler 证明了在无外力作用之下，一个质点如约束在一曲面上运动，则它必定是沿测地线运动. 另外，值得指出的是法国数学家 G. Monge 及其学派，他们对曲面论的建立也很有贡献，关于曲线和曲面理论的第一部独立的著作

《分析学在几何中的应用》是 Monge 在 1807 年出版的书，他的工作中反映出他对微分方程的兴趣. 在这些数学家的研究中，可以看到，力学、物理学与天文学以及技术与工业的日益增长的要求是促使微分几何发展的因素.

F. Frenet（1852 年）和 J. Serret（1851 年）得出了曲线的基本方程，现在通称为 Frenet 公式. 后来，G. Darboux 创造了空间曲线的活动标架概念，完整地建立起曲线理论：如果两条空间曲线 C_1 与 C_2 的曲率 k 和挠率 τ 作为曲线弧长的函数（即 $k(s)$ 与 $\tau(s)$）是相同的，则除了在空间的位置的差异之外，这两曲线 C_1 与 C_2 是完全相同的. 因而可以说曲线的所有性质都包含在它的长度、曲率、挠率的依赖关系之中. 这三个概念的研究成为研究曲线论的基础. 同时，利用它们，可以得到曲面理论中的许多概念.

在三维欧氏空间 R^3 中，与曲线相比，曲面有着重要得多的性质. 设 x_1, x_2, x_3 为 R^3 的笛氏直角坐标，则 R^3 中曲面 S 的参数方程为

$$x_i = x_i(u,v) \quad (i = 1,2,3), \tag{1}$$

曲面 S 的几何性质完全由被称为曲面的第一、第二基本形式的两个二次微分形式所决定. 第一基本形式即

$$\mathrm{d}s^2 = \mathrm{d}x \cdot \mathrm{d}x = \sum_{i=1}^{3} \mathrm{d}x_i^2 = E\mathrm{d}u^2 + 2F\mathrm{d}u\mathrm{d}v + G\mathrm{d}v^2, \tag{2}$$

E、F、G 都是 u,v 的函数. 第二基本形式是

$$\mathrm{II} = -\mathrm{d}x \cdot \mathrm{d}n = L\mathrm{d}u^2 + 2M\mathrm{d}u\mathrm{d}v + N\mathrm{d}v^2, \tag{3}$$

这里 n 是曲面 S 的单位法向量，L, M, N 都是 u,v 的函数. 曲面上一点的一个方向 $\mathrm{d}u:\mathrm{d}v$ 所对应的法曲率 $= \dfrac{L\mathrm{d}u^2 + 2M\mathrm{d}u\mathrm{d}v + N\mathrm{d}v^2}{E\mathrm{d}u^2 + 2F\mathrm{d}u\mathrm{d}v + G\mathrm{d}v^2}$.

1827 年德国数学家 C. F. Gauss 的论文“弯曲曲面的一般研究”在微分几何学的历史上有重大的意义. 微分几何发展经历了 150 年之后，Gauss 抓住了微分几何中最重要的概念和带有根本性

的内容，他在论文中建立了曲面的内在几何学，其主要思想是强调了曲面上只依赖于第一基本形式的一些性质，例如曲面上曲线的弧长，两条曲线的夹角，曲面上一区域的面积，测地线，测地曲率和总曲率，等等，这些被称为曲面的内在性质.

Gauss 之前的几何学家，用外在的方法计算曲面的总曲率，即在研究曲面时，把曲面与外围欧氏空间 R^3 相联系，找出曲面上一点的主方向，这是使法曲率取到极值的两个方向，再计算两主方向的法曲率的乘积，这便是总曲率 K，其公式是

$$K = \frac{LN - M^2}{EG - F^2}. \tag{4}$$

这是 Euler 的研究. Gauss 的方法是内在的，即由曲面的第一基本形式确定了曲面的总曲率 K，其表达式包含了 E,F,G 及其到二阶为止的偏导数，所以总曲率 K 通常也称为 Gauss 曲率，这是 Gauss 的著名发现，这两种表达式之间有着深刻的差异. Gauss 在他的论著 Theorema Egregium 中阐明了这一点，他说："如果一个弯曲的曲面可展开到任何另外的曲面上去，则每点的曲率是保持不变的"(这里，"可展"表示1对1的保持距离的映照). Gauss 建立的内在几何学有着深远的影响，是在微分几何上的一关键而重大的突破，但那时并未被人们所认识.

§2 Riemann 几何学的产生及发展

Gauss 的曲面内在几何学指出：以曲面的第一基本形式为基础，可以确定曲面的许多基本几何性质. 非欧几何的出现，表明可能的几何学并不是单一的. 在这两个重要的基础上，产生了 Riemann 几何学. Riemann 几何的出现在数学上与物理上均十分重要，具有重大而深远的意义和影响.

1854 年德国数学家 B. Riemann 就任哥丁根大学首席教授时

发表了题为《论几何学的基础假设》的就职演讲，Riemann 将曲面本身看成一个独立的几何实体，而不是把它仅仅看作欧氏空间的一个几何实体. 他发展了空间的概念，首先提出了 n 维微分流形（当时称为多重广延量）的概念，其中的点用 n 个实数 $(x^1, x^2, \cdots, x^n)$ 作为坐标来描述，曲线弧 $x^i = x^i(t)(i = 1, 2, \cdots, n; a \leqslant t \leqslant b)$ 的长度由公式

$$s = \int_a^b \sqrt{\sum_{i,j=1}^n g_{ij}(x(t)) \frac{\mathrm{d}x^i}{\mathrm{d}t} \frac{\mathrm{d}x^j}{\mathrm{d}t}}\, \mathrm{d}t$$

给出，这里 $g_{ij}(x)$ 是坐标的函数，关于 i, j 对称，并且二次微分型

$$\mathrm{d}s^2 = \sum_{i,j=1}^n g_{ij}(x)\mathrm{d}x^i\mathrm{d}x^j \tag{5}$$

是正定的. (5) 称为 Riemann 度量，Riemann 几何的基本出发点是 Riemann 度量.

Riemann 认识到度量(5) 是加到流形上去的一个结构. 因此，同一流形可以有众多的 Riemann 度量. Riemann 以前的几何学家只知道曲面 S 的外围空间 R^3 的度量赋予曲面 S 的诱导度量

$$\mathrm{d}s^2 = E\mathrm{d}u^2 + 2F\mathrm{d}u\mathrm{d}v + G\mathrm{d}v^2$$

（即第一基本形式），并未认识到曲面 S 还可以独立于 R^3 而定义，可以独立地赋予度量结构. Riemann 意识到这件事是非凡的重要，他把诱导度量与独立的 Riemann 度量两者区分开来，从而开创了以(5) 为出发点的 Riemann 几何. 这种几何以各种非欧几何作为其特例. 例如，以

$$\mathrm{d}s^2 = \frac{\sum_{i=1}^n (\mathrm{d}x^i)^2}{\left(1 + \frac{\alpha}{4} \sum_{i=1}^n (x^i)^2\right)^2} \tag{6}$$

作为两个无限邻近点的距离，当 $\alpha > 0$ 时，就是球面几何（又称为正常曲率空间的几何），$\alpha = 0$ 时就是欧氏几何，$\alpha < 0$ 时就是

Лобачевский 几何(Лобачевский 几何又称为负常曲率空间的几何或双曲几何).

如同曲面的内在几何学一样，流形上有了 Riemann 度量后就可以确定切向量的交角，各种 k 维区域的 $(1 \leqslant k \leqslant n)$ 的面积($k = 1$ 时为弧长，$k = n$ 时为体积)，此外，还可确定流形上的测地线和 Riemann 联络. 测地线是由弧长的变分所决定. 在弧长参数下，它的方程是

$$\frac{d^2x^i}{ds^2} + \Gamma^i_{jk}(x)\frac{dx^j}{ds}\frac{dx^k}{ds} = 0, ^* \tag{7}$$

式中 $\Gamma^i_{jk}(x)$ 有表达式：

$$\Gamma^i_{jk} = \frac{1}{2}g^{il}\left(\frac{\partial g_{lk}}{\partial x^j} + \frac{\partial g_{jl}}{\partial x^k} - \frac{\partial g_{jk}}{\partial x^l}\right), \tag{8}$$

称为 Riemann 联络，又称为克氏记号.

测地线是直线的推广，但在 Riemann 流形上联结二点的测地线未必总存在，也未必只有一条. 但当局限在一个充分小的适当邻域中时，联结二点的测地线总存在且只有一条.

从 Riemann 联络可以确定 Riemann 流形的曲率张量

$$R^i_{ljk} = \frac{\partial \Gamma^i_{lk}}{\partial x^j} - \frac{\partial \Gamma^i_{lj}}{\partial x^k} + (\Gamma^i_{mj}\Gamma^m_{lk} - \Gamma^i_{mk}\Gamma^m_{lj})\ (i,j,k,l,m = 1,2,\cdots,m), \tag{9}$$

它是用以表示 Riemann 流形在一点的弯曲程度的.

(i) 当 $R_{ijkl} = 0$ 时，流形局部同构于欧氏空间，也就是说，在每点 都有一邻域，它和 n 维欧氏空间的一个区域之间不仅拓扑上同胚，而且一切几何性质都相同.

(ii) 当 $R_{ijkl} = K(g_{ik}g_{jl} - g_{il}g_{jk})$($K =$ 常数)时，流形局部地和常曲率($= K$)的空间同构. 标准的单连通完备的正常曲率 K 的空

---

* 上下重复的指标表示作和，如(7)左边第二项即为 $\sum_{j,k=1}^{n}\Gamma^i_{jk}(x)\frac{dx^j}{ds}\frac{dx^k}{ds}$.

间是半径为$\sqrt{\frac{1}{K}}$的n维球面.标准的单连通、完备的负常曲率空间是 Лобачевский 空间,它们的度量在适当坐标下均可变为(6)的形式.

Riemann 几何学的早期所研究的实际上是 Riemann 流形的局部性质,即一个充分小的邻域中的性质.在20世纪中,Riemann 几何整体化了,它所研究的对象是添上 Riemann 度量g的微分流形M,记为(M,g).n维微分流形是一种拓扑空间,它为一系列邻域$\{U_A\}$(有限或可列无限)所覆盖,每一邻域U_A和n维欧氏空间R^n的一个区域相同胚,称为流形的一页地图,两个邻域的公共部分之间有坐标变换,变换公式是双向$C^k(1\leqslant k\leqslant\infty)$的.度量$g$是指在每一页地图中都有给定的 Riemann 度量,当然,在两个邻域的公共部分,Riemann 度量是相符的,即若$U_A\cap U_B$非空,依U_A上所给的度量和U_B上所给的度量是相符的,用它们计算$U_A\cap U_B$中曲线的弧长,其结果是相同的,也就是

$$\underset{A}{g_{\alpha\beta}}\mathrm{d}\underset{A}{x^\alpha}\mathrm{d}\underset{A}{x^\beta}=\underset{B}{g_{\alpha\beta}}\mathrm{d}\underset{B}{x^\alpha}\mathrm{d}\underset{B}{x^\beta}\quad(\alpha,\beta=1,\cdots,n).$$

关于整体 Riemann 几何的下述一些带有普遍性的基本事项是已知的:

(1) 任何一微分流形都能带有 Riemann 度量.

(2) 对任何两点A,B可定义其距离为

$$\mathrm{d}_{AB}=\inf l_{\widehat{AB}},$$

其中$l_{\widehat{AB}}$是联结A和B的任一分段C^1的曲线弧的长度,这样,Riemann 流形就成为距离空间.当这个距离空间完备时,就称为完备的 Riemann 流形.Riemann 流形为完备的充要条件是:它的任一测地线可以依弧长无限延伸(也可以是闭曲线).紧致无边的 Riemann 流形必为完备的.

(3) 任何一n维 Riemann 流形(M,g)必可以作为某一N维欧氏空间R^N中的子流形来实现,也就是:在R^N中有一n维子流形

(V_n,h)，其度量 h 是按照 R^N 的弧长来定义的，并且 (V_n,h) 和 (M,g) 有一个同胚关系，使对应的曲线弧的弧长完全相同．当 M 是紧致时，R^N 的维数 N 可取为 $\frac{1}{2}n(3n+11)$，当 M 是非紧致时，它可取为 $\frac{1}{2}n(n+1)(3n+11)$．N 这个数值在种种特殊情形下是可以降低的，例如对球 S^n 而言，$N=n+1$，但在一般情况下，N 的最低值尚不清楚．

Riemann 几何学的古典工具是张量分析，随着 Riemann 几何学的整体化，张量分析也改变了面貌，而且外微分形式也成为一项重要的工具，微分流形 M 的每点 P 可以定义该点的切向量，如 P 属于一坐标区域 U，局部坐标系为 $(x^1,\cdots,x^n)$，该点的切向量由一阶齐次的微分算子（即方向求导的算子）

$$X=v^i\frac{\partial}{\partial x^i} \tag{10}$$

所定义，点 P 的切向量全体构成一个向量空间 T_PM，其维数为 $n(=\dim M)$．在每点均给一向量，则得到向量场，T_PM 的对偶空间称为余切空间，记为 T_P^*M，其中元素称为协变向量，每点给一协变向量，就得协变向量场．点 P 的 p 阶反变、q 阶协变的张量就是关于 T_P^*M 上 p 重线性、关于 T_PM 上 q 重线性的函数，例如二阶协变张量 K 就由

$$\begin{aligned}K(X,Y)&=K\left(v^i\frac{\partial}{\partial x^i},w^j\frac{\partial}{\partial x^j}\right)\\&=K\left(\frac{\partial}{\partial x^i},\frac{\partial}{\partial x^j}\right)v^iw^j=K_{ij}v^iw^j\quad\left(Y=w^j\frac{\partial}{\partial x^j}\right)\end{aligned}$$

所确定．每点均给一张量就得到张量场．向量场、张量场等一般都假定是可微分的，Riemann 度量就是一个对称的二阶协变张量场，曲率张量是一阶反变、三阶协变的张量场，而由 $R_{ijkl}=g_{ih}R^h_{jkl}$ 所定义的曲率张量是四阶协变的，常记

$$R\left(\frac{\partial}{\partial x^i},\frac{\partial}{\partial x^j},\frac{\partial}{\partial x^k},\frac{\partial}{\partial x^l}\right)=R_{ijkl}. \tag{11}$$

在 T_pM 上，$\frac{\partial}{\partial x^i}(i = 1,2,\cdots,n)$ 是一组基，T_p^*M 上相应的对偶基记为 dx^i，即 $dx^i\left(\frac{\partial}{\partial x^j}\right) = \delta^i_j$，$T_p^*M$ 的元素可表示为 $a_i dx^i$，而协变向量场在一个坐标区域中可表示为 $a_i(x)dx^i$. 协变向量场又称为一次外微分形式，r 阶反称的协变张量场又称为 r 次的外微分形式，在一个坐标区域中可记为

$$\Omega = \frac{1}{r!}a_{i_1\cdots i_r}dx^{i_1} \wedge \cdots \wedge dx^{i_r} \tag{12}$$

($a_{i_1\cdots i_r}$ 关于下标为反称). 对外微分形式，可进行外微分运算 d：

$$d\Omega = \frac{1}{(r+1)!}\left(\frac{\partial a_{i_1\cdots i_r}}{\partial x^{i_0}} - \frac{\partial a_{i_0 i_2\cdots i_r}}{\partial x^{i_1}} + \cdots + (-1)^r \frac{\partial a_{i_0\cdots i_{r-1}}}{\partial x^{i_r}}\right) \cdot dx^{i_0} \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_r}, \tag{13}$$

并成立著名的 Poincaré 引理：$d^2\Omega = 0$ 对任何外微分形式 Ω 成立. 流形上 r 次外微分形式的全体构成一个线性空间，记为 $\wedge_r$，$\wedge_0$ 就是函数的全体，$\wedge_1$ 就是一次外微分形式的全体，$\wedge_r(r > n)$ 是平凡的(只有零元素). 把外微分 d 看成 $\wedge_r$ 到 $\wedge_{r+1}$ 的一个线性映照，其核空间为

$$Z_r = \{\Omega \in \wedge_r : d\Omega = 0\}, \tag{14}$$

其中元素称为闭形式，又记 d：$\wedge_{r-1} \to \wedge_r$ 的象空间为

$$B_r = \{\Omega \in \wedge_r : \Omega = d\pi, \pi \in \wedge_{r-1}\}, \tag{15}$$

其中元素称为正合形式. 由 Poincaré 引理，$B_r \subset Z_r$ 令 $H^r = Z_r/B_r$，其中每一元素称为 r 次的上同调类. 对于紧致流形 M，有著名的 de Rham 定理：H^r 是有限维的，其维数即为 M 的第 r 个 Betti 数. 这个定 理表明，流形的某些拓扑性质可由其上的外微分形式的集合所定出，这是一个有重要作用的定理，它建立了微分结构与拓扑结构的联系，利用 Riemann 度量，还可以引入调和形式的概念，W. V. D. Hodge 证明，在每一上同调类中，有且只有一个(除常数因子外)调和形式. 线性无关的 r 阶调和形式的个数就是流形 M 的第 r 个 Betti 数. 这个定理可以用来计算 Betti 数.

我们还应特别指出，Riemann 几何学的发展和广义相对论的产生有非常密切的关系，A. Einstein 在完成狭义相对论之后，进而研究引力理论，他发现，Riemann 几何是表述他的引力理论的最好的数学工具. 他的理论认为：有引力场的 4 维时空，带有一个度量

$$ds^2 = g_{ij}dx^i dx^j, \tag{16}$$

但这个度量不是正定的二次微分型，而是不定的二次微分型，更具体地说，对时空中的每一点，把这二次型通过坐标变换对角化后，其符号是(＋＋＋－). 这里的 g_{ij} 称为引力势. 是由 Einsteim 方程

$$R_{ij} - \frac{R}{2}g_{ij} = \frac{k}{c^2}T_{ij} \tag{17}$$

和一定的初始条件(或边界条件)来决定的. T_{ij} 是描述产生引力场的物质的能量动量张量，k 是引力常数，c 为光速.

在引力场中，试验质点的运动方程就是这个度量的测地线方程

$$\frac{d^2x^i}{ds^2} + \Gamma^i_{jk}\frac{dx^j}{ds}\frac{dx^k}{ds} = 0 \quad (i,j,k = 1,2,3,4), \tag{18}$$

因而克氏记号 Γ^i_{jk} 起着引力加速度的作用.

在真空情形下，Einstein 方程化为

$$R_{ij} = 0, \tag{19}$$

它的球对称解(在适当边界条件下)称为 Schwarzschild 解，表示了球状星体周围的引力场. 又 Kerr 解是描述真空引力场的一种轴对称解，它相应于转动星体所产生的引力场.

Einstein 引力理论发表后，立即得到天文学上的验证，从而也推动了 Riemann 几何学的迅速发展. 微分几何的整体化对于具非正定度量的流形的研究也起了推进作用，因为它符合于大尺度时空结构和宇宙学研究的需要. 应用这种整体化的技巧，可以解释黑洞的边界并非时空的实质性奇点，又能证明各种时空奇点的存在性. 度量符号为(＋，…，＋，－)的流形称为 Lorentz 流形，其几何学近年来在迅速发展.

§ 3　Riemann 几何学中的几个重要方向

Riemann 几何发展到现在，已成为微分几何的核心部分，既有极为丰富的内容，又有广泛的应用，研究的"热点"也很多，我们在这里择要简述其中的一部分.

（1）曲率的拓扑

一个微分流形可以带有无限多个 Riemann 度量，但是流形的拓扑性质对它所可能带有的 Riemann 度量有着强烈的制约性，Riemann 流形的曲率是微分几何中最重要的几何量之一，曲率和流形拓扑结构之间的关系是一个十分重要的问题.

这方面最早的结果是有关定向闭曲面的 Gauss-Bennet 定理：

$$\frac{1}{2\pi}\int K\mathrm{d}s = \chi. \tag{20}$$

这里 $\chi = 2(1-g)$ 是曲面的 Euler 数，其中 g 是曲面的洞的个数（又称为亏格）. χ 显然是拓扑不变量. 这式子表示了高斯曲率 K 和曲面拓扑性质的联系（例如，有洞的闭曲面不能带有 Gauss 曲率处处为正的 Riemann 度量等）. 40 年代，C. B. Allendoerfer 和陈省身用不同的方法把这个公式推广到偶数维的紧致 Riemann 流形上去，这是很重要的一项进展.

另一个早期的重要结果是 J. Hadamard 和 E. Cartan 的定理：单连通，截面曲率非正的 n 维完备的 Riemann 流形必同胚于 n 维欧氏空间 R^n. 由此可以知道，对于单连通的紧致流形所带的 Riemann 度量，其截面曲率不能处处为非正.

对于 Riemann 流形来说，有三种不同层次的曲率，即截面曲率，Ricci 曲率和数量曲率.

截面曲率是相应于每点某一平面方向的曲率. 设这平面由两

个单位切向量 ξ_1 和 ξ_2 所张成，则截面曲率为

$$K(\xi_1,\xi_2) = \frac{R_{ijkl}\xi_1^i\xi_2^j\xi_1^k\xi_2^l}{(g_{ik}g_{jl} - g_{il}g_{jk})\xi_1^i\xi_2^j\xi_1^k\xi_2^l}, \tag{21}$$

所以截面曲率是点的位置以及平面方向的函数.

截面曲率有时会对流形作出很强的限制，例如，设 M 是 n 维紧致单连通 Riemann 流形，若其截面曲率 K_M 满足 $\frac{1}{4} < K_M \leqslant 1$，那么 M 和球面同胚，这类定理称为挤压(Pinching)定理，是由 Berger 和 Klingenberg 所得到. 最近 M. Micallef 和 J. D. Morre 推广了这个定理.

Ricci 曲率张量为

$$R_{ij} = g^{kl}R_{kijl}, \tag{22}$$

对于一个单位向量 ξ，Ricci 曲率之值为

$$\mathrm{Ric}(\xi) = R_{ij}\xi^i\xi^j, \tag{23}$$

数量曲率 S 为

$$S = g^{ij}R_{ij}. \tag{24}$$

Ricci 曲率和流形拓扑性质之间的关系已有许多结果，例如，S. B. Myers 定理说，当流形的 Ricci 曲率处处都不小于一个正常数时，流形必是紧致流形. 又如 A. Lichnerowicz 证明了：如果紧致 Riemann 流形的 Ricci 张量是正定时，则第一 Betti 数为 0.

截面曲率恒为正常数的流形已有相当完整的分类定理，对于截面曲率恒为负常数的流形，情况尚未完全弄清.

关于曲率和流形拓扑之间的关系，目前仍然是微分几何主要的研究方向之一，存在着大量的尚未解决的问题，例如可参见丘成桐所收集的问题集.

(2) 子流形，极小子流形

浸入于 $n+p$ 维 Riemann 流形 (M^{n+p},g) 中的 n 维子流形 M^n，依照 M^{n+p} 的度量(即计算曲线弧长的方法)，可以诱导出 M^n 的

Riemann 度量 g_1. Riemann 空间子流形(M^n, g_1)理论是普通微分几何中曲面论的推广. 当 $M^n(n > 2)$ 的余维数为 1(即 $p = 1$) 时, M^n 称为超曲面. 除第一基本形式(即度量 g_1) 外, 子流形还有第二基本形式; 除子流形的内在几何学外, 还要研究它在外围流形(M^{n+p}, g)中所体现出来的种种性质. 例如, 当(M^n, g_1)的任何测地线都是(M^{n+p}, g)的测地线时, M^n 就称为(M^{n+p}, g)的全测地曲面, 它是欧氏空间中平面的扩充. 已经知道, 如果过 M^{n+p} 每点且和任意一个 n 维平面方向相切的 n 维全测地曲面都存在的话, 那末(M^{n+p}, g)一定是常曲率的 Riemann 流形.

在各种子流形中, 极小子流形具有特殊的重要性. 平均曲率为 0 的曲面(或子流形) 通称为极小曲面(或极小子流形). 1847 年比利时物理学家 J. Plateau 提出了如下著名问题: 在欧氏空间 R^3 中已给一闭曲线 c, 求一以 c 为边界的曲面, 使其面积为最小. 这种曲面一定是极小曲面, 这问题被称为 Plateau 问题. 虽然这个问题可以通过物理实验来处理(利用肥皂膜的表面张力使其面积缩为最小), 但数学上处理却极为困难. 经过了近百年的研究, 动用了复变函数, 偏微分方程, 变分学等多种领域的深刻工具, 到 1930 − 1931 年, 才由 T. Rado 和 J. Douglas 予以解决. 在其后, 极小曲面的研究还在继续, 近年来又有较大的进展. 在古典理论中, 过去人们只知道正螺旋面, 悬链面和平面这三种具有限拓扑型(即同胚于除去有限点后的闭曲面) 的极小曲面, 但 80 年代以来, J. Hoffmann, D. Hoffmann 与 W. Meeks 借助于计算机发现了无限个具有限拓扑型的极小曲面, 并证明了一系列值得注意的事实, 引起了极大的兴趣. 又谷超豪发现, 每个极小曲面都对应于一个三维不可压缩流体的有奇性的流场.

近年来高维欧氏空间的极小超曲面受到很大的注意, 在局部范围内, 超曲面的方程可表示为

$$x^n = f(x^1, \cdots, x^{n-1}), \tag{25}$$

极小超曲面满足微分方程

$$\sum_{i=1}^{n-1} \frac{\partial}{\partial x^i}\left(\frac{\partial f}{\partial x^i} / \sqrt{1 + |\nabla f|)^2|}\right) = 0, \tag{26}$$

这就是面积积分的 Euler 方程.

1915 年,S. Bernstein 证明:$n = 3$ 时,定义在全平面上的极小曲面方程的解只有平面. 在 1965 — 1968 年间,E. De Giorgi,F. J. Almgren 和 J. Simons 证明:$n \leqslant 8$ 时,类似的 Bernstein 定理也是成立的. 非常有趣的是:E. Bomberi,E. De. Giorgi 和 E. Giusti 证明:在 $n \geqslant 9$ 时,存在定义在整个平面 R^{n-1} 上的满足上述极小超曲面方程的非平面解.

对于球面中的极小超曲面,1969 年陈省身提出了所谓"球面 Bernstein 问题":S^{n+1} 中的与 S^n 微分同胚的极小超曲面是否为全测地的?在 $n = 2$ 的情形,F. J. Almgrem,E. Calabi 分别证明了这个结论成立. 然而,项武义运用 Lie 群表示理论成功地构造了 $S^{n+1}(n = 3,4,5,6,7,9,11,13)$ 中无限多个非全测地的与 S^n 微分同胚的极小超曲面的例子. 关于球面中极小子流形的数量曲率的研究在最近一段时期也有了很大进展. 在这一方向上,J. Simons 首先证明了如下的挤压(Pinching)定理:设 M^n 为半径为 1 的 $n + p$ 维球面 $S^{n+p}(1)$ 中的紧致极小子流形,如果 M^n 的数量曲率 S 满足 $S > n\left[n - 1 - \left(2 - \frac{1}{p}\right)^{-1}\right]$,则 $S = n(n-1)$. 这说明数量曲率有一个空隙. 随后,陈省身,do Carmo,S. Kobayashi 决定了数量曲率 $S = n\left[n - 1 - \left(2 - \frac{1}{p}\right)^{-1}\right]$的所有极小子流形. 另外,滕楚莲与彭家贵证明了 S 的第二空隙的存在性.

常曲率空间形式中主曲率等于常数的超曲面称为等参超曲面,这是由 E. Cartan 首先引入的的. 球面 S^n 的等参超曲面的分类只有部分的结果,有一类是利用 Clifford 代数的表示制作出来的. 其中包含了许多极小超曲面. 王启明给出了这种极小超曲面的相

当完全的分类，并研究了它们的拓扑性质，作为一项推论，他举出了 $S^{199}(1)$ 中两个紧致极小超曲面，它们的数量曲率都是常数38412，它们是微分同胚的，但并不等距，这就解决了陈省身在1968年所提出的一个问题：S^{n+1} 中两个具有相同常数数量曲率的紧致极小超曲面在 S^{n+1} 中是否等距的？陈省身和合作者在1978年又重提了这个问题. 他曾猜想答案是肯定的，但王启明举出了反例，王启明还和丘成桐合作，解决了极小超曲面的一些其他问题，如在 R^{16} 中作出了一个稳定的、非极小化的光滑超曲面.

极小子流形理论的应用也非常广泛，在1980年左右，R. Scheon 和丘成桐使用极小曲面理论证明了广义相对论中著名的正质量猜测：在一个孤立物理系统中，能量动量向量必为指向未来的类时向量，但 Minkowski 时空除外.

最近，R. Schoen 和丘成桐又推广了他们的极小曲面论证方法，用来证明广义相对论中的一个奇点定理. 这个定理的大意是：质量高度集中必然导致时空奇点的存在. 另外，运用极小子流形理论，数学家们发现了流形的几何与拓扑的很多奥秘. 譬如，R. Scheon 和丘成桐证明了具有正的 Ricci 曲率的非紧三维流形必与三维欧氏空间微分同胚等.

Minkowski 空间 R^{n+1} 是平坦的 Lorentz 流形，$n=3$ 时，它刻划相对论中的时空. R^{n+1} 的度量可写成

$$ds^2 = dx_1^2 + \cdots + dx_n^2 - dx_{n+1}^2 \quad (-\infty < x_i < \infty). \tag{27}$$

由于度量非正定，所以情况显得复杂一些. 70年代以来，Minkowski 空间 R^{n+1} 的极值曲面（即面积变分为0或平均曲率为0的曲面）已经有了不少研究，在这种空间中，曲面有类空（度量为正定的），类时（度量为非退化、非正定的）和类光（度量为退化的）之分. 郑绍远和丘成桐研究了类空的“极大超曲面”的存在性问题. 近年来，谷超豪开始了混合型极值曲面的研究，这种曲面既包含类空部分，又包含类时部分，因而需要求解拟线性混合型偏微分

方程，他得出了极值曲面的一般表达式，并且给出了从 R^n 中解析曲线出发作出整体的极值曲面的显式方法. 另外，他还解出了一类边值问题.

Minkowski 空间的曲面理论和欧氏空间曲面论有许多实质上不同之处，欧氏空间 R^3 中的球面是刚性的，即和球面相等距的曲面必互相合同，Hano、Nomizu 和胡和生用不同的方法证明了嵌入于 Minkowski 空间 R^{2+1} 中的非欧平面不是刚性的，胡和生还证明，在 Minkowski 空间 R^{2+1} 中存在一系曲率为 -1 的类空曲面，它们是完备的图，$x_3 = f(x_1, x_2)(-\infty < x_1, x_2 < \infty)$，但是从度量的角度，它却是不完备的. 这个定理给出了 R^3 中完备图与 R^{2+1} 中完备图的一个本质区别.

由于极小子流形（以及另外一些具有特定性质的子流形，如平均曲率等于非零常数的曲面）在微分几何，理论物理，拓扑学，偏微分方程等领域中的重要性，对于它们的研究继续在深入中.

(3) 特征值问题

Laplace 方程在数学物理中起着十分重要的作用，真空中的静电场，经典引力场，热的稳态分布，弹性膜的平衡等都要解 Laplace 方程. 在求解热传导或膜振动动态问题时，就要求解 Laplace 方程的特征值问题，即在一定边界条件下，求

$$\triangle u = \lambda u \tag{28}$$

的非平凡解，这里 $\triangle = \frac{\partial}{\partial x_1^2} + \cdots + \frac{\partial}{\partial x_n^2}$，称为 Laplace 算子. 这个问题只有对特殊的 λ 有非平凡解，这种 λ 称为特征值. 在膜的情况下，当 $n = 2$ 时，特征值表示膜振动的固有频率（如一只鼓所发出的声音的频率）.

Laplace 算子可推广到 Riemann 流形上，它是微分几何中最重要的微分算子之一，在局部坐标下其表达式为

$$\triangle_M = g^{ij}\left(\frac{\partial^2}{\partial x^i \partial x^j} - \Gamma_{ij}^k \frac{\partial}{\partial x^k}\right). \tag{29}$$

当 Riemann 流形是紧致的情形，它的 Laplace 算子有离散的特征值：$0 \leqslant \lambda_1 \leqslant \cdots \leqslant \lambda_i \leqslant \cdots$. 在近几十年里，关于特征值方面的研究非常活跃. A. Lichnerowicz 在 1958 年首创了对 Riemann 流形的第一特征值的估计：若 M^n 为完备 Riemann 流形，又它的 Ricci 曲率满足 $\mathrm{Ric}_M \geqslant (n-1)c > 0$，则 $\lambda_1 \geqslant nc$，这里 c 是正常数，接着 M. Obata 证明了当 $\lambda_1 = nc$ 时，M^n 就是曲率为 c 的球面 $S^n(c)$. 1975 年，郑绍远给出了 λ_1 的上界，这一上界只依赖于流形的直径及 Ricci 曲率的下界. 后来 P. Li 和丘成桐运用“梯度估计”方法得到 λ_1 相应的下界，钟家庆和杨洪苍在 1983 年改进了 P. Li 和丘成桐的结果，得到下界的最佳估计.

估计特征值的空隙对于微分几何与理论物理都是很有意义的. 1985 年，M. Singer，B. Wang，丘成桐及丘成栋在 R^n 的凸域 Ω 上，关于 Dirichlet 边界条件的开头二个特征值之间的空隙作了估计，他们证明了：$\lambda_2 - \lambda_1 \geqslant 4\pi^2/d(\Omega)^2$，这里 $d(\Omega)$ 是区域 Ω 的直径. 1986 年钟家庆和于际黄把上述下界改进到 $\pi^2/d(\Omega)^2$. 最近，A. L. Wang 将上述结果推广到球面中的凸区域. 我们相信，在今后几年中，关于特征值的空隙会有更多精彩的结果.

特征值研究的另一课题是所谓谱几何：一个 Riemann 流形的全体特征值能够反映出这流形的多少几何信息？首先，通过研究热核的迹 $\Sigma e^{-\lambda_i t}$，人们知道，当 $t \to 0$ 时，从这无穷级数的渐近行为可以得到体积及总数量曲率等；有一些特殊流形的结构可以完全由它的谱所决定，譬如，维数不大于 6 的球面就是如此，然而，在一般情况不能指望流形的所有几何量可以由它的谱所描述. 事实上，在 1984 年，C. Gordon 及 Wilson 发现了在某个紧致流形上存在具有相同谱的连续度量族. 现在数学家们还未找到 Riemann 流形的结构能由其谱所决定的充要条件.

近些年来，特征值的应用越来越广泛，这里，我们仅举几个例子. 1975 年，郑绍远用特征值比较定理建立了下面的直径定理：设 M^n 为完备 Rimann 流形，$\mathrm{Ric}_M \geqslant (n-1)c > 0$，如 $d(M) = \pi/\sqrt{c}$，则 M^n 与 $S^n(c)$ 等距，根据 H. I. Choi 与 A. N. Wang 所得到的 $S^{n+1}(1)$ 中紧致嵌入极小超曲面的第一特征值 $\lambda_1 \geqslant \frac{n}{2}$ 的结果以及丘成桐和 P. C. Yang 所建立的紧致 Riemann 面的第一特征值和其体积的关系，人们知道：$S^3(1)$ 中紧致嵌入极小曲面的体积有一个仅依赖于它的亏格的上界. 1984 年，郑绍远、P. Li 和丘成桐通过热核的迹的上界估计得到了球面中极小子流形体积的空隙性. 这一空隙性后来被林峻岷与夏昌玉用来研究球面中偶数维极小子流形的数量曲率的积分估计. 另外根据 A. Ros 和 S. Montiel 的工作，我们知道，如果丘成桐的猜想"$S^{n+1}(1)$ 中紧致嵌入极小超曲面的第一特征值为 n"成立，则 Lawson 的猜想"$S^3(1)$ 紧致嵌入极小环面为 Clifford 环面"也是成立的. 这里，Clifford 环面是指 $S^3(1): x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1$ 中的曲面

$$x_1 = \frac{1}{\sqrt{2}}\cos\theta, x_2 = \frac{1}{\sqrt{2}}\sin\theta, x_3 = \frac{1}{\sqrt{2}}\cos\varphi, x_4 = \frac{1}{\sqrt{2}}\sin\varphi.$$

因而，特征值问题这一课题除了本身包含着丰富的内容和技巧外，它在 Riemann 流形的几何与分析中的应用也很广泛.

(4) 调和映照

调和映照是微分几何中重要的研究对象，它与数学物理有密切的联系.

先介绍定义

设 M 与 N 为 Riemann 流形或 Lorentz 流形，$\varphi: M \to N$ 是 C^2(或 L_1^2) 映照. 映照 φ 的能量积分由

$$E(\varphi) = \int_M e(\varphi) \mathrm{d}V_M \tag{30}$$

所定义，$e(\varphi)$ 是能量密度，在局部坐标下它有如下的表达式

$$e(\varphi)=\underset{N}{g_{\alpha\beta}}(\varphi)\frac{\partial\varphi^{\alpha}}{\partial x^{i}}\frac{\partial\varphi^{\beta}}{\partial x^{j}}\underset{M}{g^{ij}}(x)\quad(i,j=1,\cdots,m;\alpha,\beta=1,\cdots,n),\tag{31}$$

这里$\underset{M}{g^{ij}}$与$\underset{N}{g_{\alpha\beta}}$分别是流形 M 与 N 的度量张量的反变分量与协变分量. $n=\dim N, m=\dim M$.

如果映照 φ 是 $E(\varphi)$ 的临界点，则称 φ 为 $M\to N$ 的调和映照. 其微分方程为

$$\underset{M}{g^{ij}}\left(\frac{\partial^{2}\varphi^{\alpha}}{\partial x^{i}\partial x^{j}}-\underset{M}{\Gamma^{k}_{ij}}\frac{\partial\varphi^{\alpha}}{\partial x^{k}}+\underset{N}{\Gamma^{\alpha}_{\beta\gamma}}\frac{\partial\varphi^{\beta}}{\partial x^{i}}\frac{\partial\varphi^{\gamma}}{\partial x^{j}}\right)=0,\tag{32}$$

这里$\underset{M}{\Gamma^{k}_{ij}}$ 和$\underset{N}{\Gamma^{\alpha}_{\beta\gamma}}$ 分别是 M 与 N 的克氏记号，如果 M 是 Riemann 流形，则(32) 是一系非线性椭圆型方程组. 如果 M 为 Lorentz 流形，则(32) 是一系非线性双曲型方程组.

我们通过下面的一些简例来看调和映照在数学上与物理上的重要性.

① 设 $M=R$，而 N 为任意 Riemann 流形或 Lorentz 流形，则调和映照方程(32) 变为测地线所满足的微分方程. 设 $M=S^{1}$，则 $\varphi(x)$ 是闭测地线.

② 设 $N=R$，M 为任意 Riemann 流形，则调和映照方程就是流形 M 上的 Laplace 方程，而 φ 则是 M 上的调和函数.

③ 如果调和映照中，$M\to N$ 又是等距浸入，它就是一个极小浸入.

④Kähler 流形间的全纯映照是调和映照.

从下面的例子进一步看出调和映照在物理上也很重要.

⑤ 非线性 σ- 模型或称手征场是一种引起广泛注意的非线性场论，它实际上是 $M\to N$ 的调和映照，这里 M 是 Minkowski 时空或欧氏空间，而 N 通常是 Lie 群或齐性 Riemann 空间.

⑥ 粒子物理学中有时用弦来作为强子的模型，在弯曲时空

(Lorentz 流形)V 中,描述弦运动的世界面是一类时的二维曲面,它的运动方程可写为

$$\varphi^{\alpha}_{\tau\tau}-\varphi^{\alpha}_{\sigma\sigma}+\underset{V}{\Gamma^{\alpha}_{\beta\gamma}}(\varphi^{\beta}_{\tau}\varphi^{\gamma}_{\tau}-\varphi^{\beta}_{\sigma}\varphi^{\gamma}_{\sigma})=0, \tag{33}$$

这实际上是二维 Minkowski 平面 $R^{1,1}\to V$ 的调和映照方程.

调和映照的一个基本问题是在映照的每一个同伦类中调和映照的存在性(包括显式作法)问题,迄今已有许多重要结果,现举出一部分.

$M=S^1$,调和映照是闭测地线,当 N 为紧致时,Hilbert 的经典定理已给出了回答:每一闭曲线的同伦类中,必有一闭测地线.

M 为 2 维紧致流形,若 N 为紧致,且其第二同伦群 $\pi_2(N)=0$,则回答也是肯定的.

当 M,N 均为紧致,$\dim M\geqslant 2$ 时,有下列的结论:当 N 的截面曲率 $\mathrm{Riem}N\leqslant 0$ 时,M 到 N 的映照的每一同伦类中必存在一个调和映照,使能量在该同伦类中达到极小值,这是 J. Eells 和 J. H. Sampson 在 1964 年所得到的一个重要的存在定理.

如 M 有边,可提 Dirichlet 问题. R. Hamilton 证明了:在 $\varphi:M\to N,\varphi|_{\partial M}=h$ 的每一同伦类中,必存在使能量达到极小值的调和映照. 这个定理可以推广到出现 $\mathrm{Riem}N>0$ 的情形,但 h 要充分小.

对于 $M\to N$ 的映照的每一同伦类 π,必存在映照序列 $\{\varphi_n\}$($\varphi_n\in\pi$),使 $E(\varphi_n)\to\inf\limits_{\varphi\in\pi}E(\varphi)$,而且 $\{\varphi_n\}$ 在 Sobolev 空间 $L^2_1(M,N)$ 中弱敛于 $\tilde{\varphi}$,如果 $\tilde{\varphi}$ 无奇点,就是所需要的调和映照. 这种正则性的证明往往是很复杂的. 1982 年 R. Schoen 和 K. Uhlenbeck 在一般情况下证明了下述的部分正则性定理:设 M,N 为紧致流形,则 $\tilde{\varphi}$ 的奇点集的 Hausdorff 维数不超过 $\dim. M-3$. 从这个定理还可以得出:如果所有 j 维球面 $S^j(j=2,\cdots,m-1)$ 到 N 的调和映照都只是常映照,则 $\tilde{\varphi}$ 必为光滑的. 这说明,从某些不存在性的结果可

以推出存在性来，又当目标流形为不可约齐性紧致 Riemann 流形时，忻元龙改进了上述定理，导出了 $\tilde{\varphi}$ 的正则性条件，它依赖于流形的维数，目标流形截面曲率的上界，数量曲率以及 Laplace 算子的第一特征值.

调和映照的显式作法，近年来发展很快，特别是 S^2 到射影空间，Grassman 流形的调和映照，已解决得很好，在有关的研究中，利用 Twistor 于制作调和映照，显得很有成效.

当 M 或 N 为 Lorentz 流形时，关于调和映照的存在性问题的研究还不多，谷超豪研究了 Minkowski 平面 R^{1+1} 到完备 Riemann 流形 N 的调和映照. 这时主要的问题是 Cauchy 问题. 他证明了整体解的存在性，由此得出的一个物理推论是：R^{1+1} 上的非线性 σ 模型是不会产生奇性的.

对于 $R^{1+n}(n>1)$ 到 N 的调和映照，Shatah 举例说明，在 $n\geqslant 3$ 时，对于某些初始条件，Cauchy 问题的整体解不存在. 在初始条件充分小时（即和常值映照充分接近时），从双曲型方程一般理论可推出整体解存在.

若 N 是 Lorentz 流形，M 是 R^{1+1} 时，Cauchy 问题的局部解可能在 t 为有限之处就变为非正则的. 当 N 是 R^{2+1} 中的半径为 1 的单位球 S^{1+1} 时，谷超豪分析了各种可能的初始条件，举出了大范围解存在的情况.

Riemann 流形 M 到 Riemann 流形 N 的调和映照的不存在性是指除常值映照外不存在其他调和映照. 我们这里假设 M，N 都是连通的.

在调和映照的能量为有限的条件下，R. Schoen 与丘成桐得到：设 M 为完备、非紧、Ricci 曲率 $\geqslant 0$ 的 Riemann 流形，N 为 $\mathrm{Riem}N\leqslant 0$ 的 Riemann 流形，则任何具有限能量的调和映照必为常值映照.

在能量有限时，Eells 与 Lemaire 得到：设 $\varphi: R^m\to N(m>2)$ 是

欧氏空间到任何 Riemann 空间 N 的具有限能量的调和映照，则 φ 是常值映照. 胡和生的工作减弱了有限能量这一条件，她得出：当 $\varphi: R^n \to N$ 的能量为无限，但发散不太快时，φ 也必为常值映照. 又 R^n 改为负常曲率空间 H^n 和某些共形平坦的非紧 Riemann 流形时，这一结果同样成立. 胡和生还把这种不存在性定理扩展到 Schwarzschild 解黑洞外的区域，但在黑洞上要满足一定的边界条件.

当 $M \to N$ 的调和映照的第二变分为非负，则称它为稳定的调和映照. 显然，极小映照是稳定的调和映照. 1980 年忻元龙得出：$n \geqslant 3$ 时，不存在从欧氏球面 S^n 到任何 Riemann 流形的稳定调和映照. 这个结果引起国内外许多人的注意，潘养廉、沈一兵等对此作了推广，将 S^n 改为更一般的流形，例如，潘养廉证明了紧致单连通 $\delta(n)$-pinched 流形 M^n 到任何 Riemann 流形的稳定调和映照的不存在性.

调和映照与某些重要的偏微分方程如 Sine-Gordon 方程、Sinh-Gordon 方程、Sine-Laplace 方程，Sinh-Laplace 方程等有密切联系. 胡和生等利用这些方程的解，得出 $R^{1,1}$或 R^2到各种"球面"的调和映照，并且构造了 $R^{2,1}$或 R^3中相应的常曲率曲面.

(5) **复流形**

一个 $2m$ 维的微分流形 M，如果可由一族坐标区域 $\{U_\alpha\}$ 所覆盖，在每一区域中可引入 m 个复数为坐标，对任何两个相交的坐标分别为 $(z^1, \cdots, z^m)$ 和 $(w^1, \cdots, w^m)$ 的坐标邻域 U, V，其坐标变换的关系

$$w^\alpha = f^\alpha(z^1, \cdots, z^m) \quad (\alpha = 1, 2, \cdots, m) \tag{34}$$

是 $z^1, \cdots, z^m$全纯函数，那末 M 就称为具有复结构的，具有复结构的流形就称为复流形.

最简单的复流形是 Gauss 复平面 C^1，其坐标为 $z = x +$

$\sqrt{-1}y$，添上无限远点 $z=\infty$ 后，就成为Gauss球面，它是 C^1 的紧化. m 维的复流形中，最简单的例子是复空间 C^m，由元素 $(z^1, \cdots, z^m)$ 的全体所构成，另一个例子是复 m 维射影空间 CP^m. 在 $C^{m+1}\backslash\{0\}$ 中二个元素 $(u^1, \cdots, u^{m+1})$ 和 $(v^1, \cdots, v^{m+1})$ 之间如存在关系 $u^i = \lambda v^i (i = 1, 2, \cdots, m+1, \lambda$ 为非零复数)，就称它们为等价的，这种等价类的全体就构成了 CP^m，$(u^1, \cdots, u^{m+1})$ 称为其中点的齐次坐标. 我们注意到 CP^1 就是 Gauss 球面.

m 维复流形可以带有 Riemann 度量，如果在各个复的坐标图中，度量可以写成为

$$ds^2 = h_{i\bar{j}} dz^i d\bar{z}^j \tag{35}$$

这里 $(h_{i\bar{j}})$ 构成正定的 Hermite 阵，那末这度量称为 Hermite 的，流形就称为 Hermite 流形. 又如果二次微分形式

$$\omega = h_{i\bar{j}} dz^i \wedge d\bar{z}^j \tag{36}$$

是一个闭形式，或等价地

$$\frac{\partial h_{i\bar{j}}}{\partial z^h} = \frac{\partial h_{k\bar{j}}}{\partial z^i}, \frac{\partial h_{i\bar{j}}}{\partial \bar{z}^k} = \frac{\partial h_{i\bar{k}}}{\partial \bar{z}^j}$$

成立，此流形称为 Kähler 流形，其度量称为 Kähler 度量.

在复流形上有一个(1,1)型张量场 J（称为复结构），在局部坐标下，如果 $z^i = x^i + \sqrt{-1}y^i$，则 J 为

$$J\left(\frac{\partial}{\partial x^i}\right) = \frac{\partial}{\partial y^i}, J\left(\frac{\partial}{\partial y^i}\right) = -\frac{\partial}{\partial x^i}. \tag{37}$$

在 Hermite 流形上，除了通常 Riemann 流形所具有的一些几何量外，还有一些与复结构有关的量，如全纯双截曲率 $\mathrm{HBRiem}(X, Y) = R(X, JX, Y, JY)$ 等.

复几何的研究起源于对 Riemann 面的研究. 对 Riemann 面，单值化定理给出了由亏格完全确定的一个简单分类，例如，单连通的 Riemann 面分别是球面、复平面和圆盘，在高维情形，分类问题是一个正在研究的重要课题，由 Riemann 面的分类导出的一个自

然的问题是，在怎样的几何条件下，一个高维的复流形可双全纯同胚于一个比较典型的流形，如复射影空间、复欧氏空间等. 这方面已有了若干相当好的结果. 例如，Mori 和肖荫堂、丘成桐分别解决了 Frankel 猜想，即任何具正双截曲率的紧 Kähler 流形必双全纯等价于复射影空间. 最近，Mok 又给出了具非负双截曲率的紧 Kähler 流形的万有覆盖空间的一个完全刻划.

曲率是一个流形的最重要的几何量之一. 在一定的曲率条件下的 Kähler 度量的确定，是很需要解决的问题. 例如，Calabi 猜测：紧 Kähler 流形上任一代表第一陈类的形式必是某个 Kähler 度量所相应的 Ricci 形式. 这里 Ricci 形式定义为 $-(2\pi\sqrt{-1})^{-1}R_{i\bar{j}}dz^i \wedge d\bar{z}^j$，而 $R_{i\bar{j}}$ 是 Ricci 张量，此猜想由丘成桐给出了证明.

在复流形之间，全纯映照是最基本的映照，而它又与调和映照的研究有着密切的关系. 我们知道，任何紧 Kähler 流形到 Kähler 流形的全纯映照必为调和映照，且在同伦类中能量达到最小，而另一方面，Kähler 流形间的许多调和映照又一定是全纯映照或反全纯映照. 例如，Eells 和 Wood 证明了任何紧 Riemann 面到 CP^n 的同伦类中能量达到最小的调和映照必为全纯映照或反全纯映照，最近，Burstall，Rawnsley 和 Salamon 证明了，S^2 到任一不可约 Hermite 对称空间的稳定调和映照必为全纯映照或反全纯映照. 这些结果对调和映照及其他几何问题的研究起着相当大的作用，例如，肖荫堂和丘成桐解决 Frankel 猜想时就用了调和映照的全纯性定理.

齐性复流形也具有特别的重要性，齐性空间可理解为 Lie 群 G 依其闭子群 H 所作的旁集空间 $M=G/H$. F. Klein 在 19 世纪提出的几何按群论分类的著名的 Erlangen 计划书，所讨论的实际上就是齐性空间的几何学，最重要的齐性空间是对称 Riemann 空间，这类空间的特点是它关于每点都是对称的（即关于每一点的对称都是等距变换）. E. Cartan 已经作出了对称空间的分类，它密切联

系于单纯 Lie 群的结构，特别是对合自同构（即平方为恒等变换的自同构）的决定.

齐性复流形的研究对于多复变函数有重要的作用，人们已研究了多种齐性复流形，如对称 Hermite 流形，J. Tits 引入的 D 空间 G/P（这里 G 是连通复 Lie 群，P 是其抛物子群，即包含 G 的极大可解子群的子群）和一般的紧齐性复流形等.

复流形的研究与多复变函数、代数几何、Lie 群理论和数学物理都有很密切的关系，体现了多种学科的交叉.

§4　纤维丛，联络和杨-Mills 方程

纤维丛是近代微分几何中很重要的一个概念，它是从研究一个微分流形切向量的集合（称为切丛）或一般的张量的集合（称为张量丛）所抽象出来的. 事实上，20 世纪初期，E. Cartan 在某些纤维丛上引入了"联络"的概念，对微分几何的发展起了重大作用. 陈省身教授对此作出重要贡献. 在 70 年代中，人们认识到联络就是理论物理中非常重要的规范势（又称杨-Mills 势），纤维丛和联络的作用更为令人注目了. 还应特别指出的是：物理学中的杨-Mills 方程进入数学之后，对纯数学的研究又产生了非凡的影响，可以说是 20 世纪数学史上的大事之一. 我们对这几个概念作一些概要的叙述.

(1) 纤维丛

最简单的纤维丛是乘积向量丛. 设 U 是一个微分流形，F 是 m 维向量空间，集合 $U\times F=\{(x,u)\mid x\in U,u\in F\}$ 称为以 U 为底，以 F 为纤维的乘积向量丛. 事实上，它就是把 U 上每一点（以普通的两个拓扑空间作乘积空间的方式）附上一个向量空间而成，它的拓扑性质由如下准则确定：$U\times F$ 中元素序列 $\{(x_\alpha,u_\alpha)\}$ 收敛于

(x,u) 的充要条件是 $x_n \to x$（在 M 中）和 $u_n \to u$（在 F 中）. $U \times F$ 上有一个投影算子 $\pi: U \times F \to F$，其定义是 $\pi(x,u) = x$. 另外还可定义截面. 它是映照 $\sigma: U \to U \times F$，使得 π 和 σ 的复合满足 $\pi \circ \sigma = id_U$（U 上的恒等映照），σ 可以理解为定义在 U 上取值于 F 的函数. 设 $\sigma_1, \cdots, \sigma_m$ 是 m 个截面，且在每点 x，它们都是线性无关的，那么任何一截面 $\sigma(x)$ 均可表示为

$$\sigma(x) = U^a(x)\sigma_a(x) \tag{38}$$

的形式. 这时 $\sigma_1, \cdots, \sigma_m$ 称为截面的一组基，乘积丛上必存在截面的基，例如取 F 的一组基 $e_1, \cdots, e_m$，令 $\sigma_a(x) = e_a$，就构成一组基.

一般的向量丛是由乘积向量丛拼接而成. 设微分流形 M 被一些区域所覆盖，即 $M = \bigcup U_p (p = 1, 2, \cdots)$，在 U_p 上有乘积丛 $U_p \times F_p$，这里 F_p 都同构于线性空间 F，又设 $\underset{p}{\sigma_1}, \cdots, \underset{p}{\sigma_m}$ 为 $U_p \times F_p$ 的截面的基. 设 $U_p \cap U_q$ 非空，$g_{pq}(x)$ 是定义在 $U_p \cap U_q$ 上，以 $m \times m$ 非异阵为值的函数（即 $g_{pq}: U_p \cap U_q \to GL(m)$），当 $x \in U_p \cap U_q$ 时，$U_p \times F_p$ 中的元素 $\underset{p}{u^a}(x)\underset{p}{\sigma_a}(x)$ 和 $U_q \times F_q$ 中的元素 $\underset{q}{u^a}(x)\underset{q}{\sigma_a}(x)$ 满足

$$\underset{p}{u}(x) = g_{pq}(x)\underset{q}{u}(x) \tag{39}$$

时，就把这两个元素视为等同的. $g_{pq}(x)$ 称为转换函数，借助于这种转换函数，就可以把这些乘积丛 $U_p \times F_p$ 拼接起来成为一个整体，这便构成了向量丛 E，M 称为它的底流形. 我们常假定 $g_{pq}(x)$ 是可微分的. 又为使拼接是合理的，应有

$$\begin{aligned} g_{pq}(x) &= g_{qp}^{-1}(x), \quad x \in U_p \cap U_q; \\ g_{pq}(x)g_{qr}(x) &= g_{pr}(x), \quad x \in U_p \cap U_q \cap U_r. \end{aligned} \tag{40}$$

容易看到，只有这样，才使拼接成为完全确定的. 定义在每一 $U_p \times F_p$ 上的投影算子可以扩到整个 E 上，成为 E 到 M 的投影算子 π，而向量丛往往就记为 $E \xrightarrow{\pi} M$. 又照样可以定义截面 $\sigma: M \longrightarrow E$，它满足 $\pi \circ \sigma = M$ 上的恒等映照 id_M，$\underset{p}{\sigma_1}, \cdots, \underset{p}{\sigma_m}$ 构成截面的局部基.

n 维微分流形 M 的切向量的全体构成一个向量丛，记为 TM，在坐标区域 U 中，$x_1, \cdots, x_n$ 是局部坐标，U 上的切向量场可用

$$X(x) = v^i(x)\frac{\partial}{\partial x_i} \tag{41}$$

表示，特别 $\frac{\partial}{\partial x^i}(i = 1, 2, \cdots, n)$ 就是截面的一组局部基. 如有另一坐标区域 V，$y_1, \cdots, y_n$ 是局部坐标，也可以把 $\frac{\partial}{\partial y^i}$ 看成一组局部截面基. 设 $U \cap V \neq \varnothing$，则在其中

$$\frac{\partial}{\partial x^i} = \frac{\partial y^j}{\partial x^i}\frac{\partial}{\partial y^j} \tag{42}$$

表示同一切向量，所以相应的转换函数为 $n \times n$ 非异阵 $\left(\frac{\partial y^j(x)}{\partial x^i}\right)$. 切丛未必是乘积丛，例如普通的球面 S^2 的切丛就不是乘积丛，因为我们已经知道 S^2 上不存在非零的连续的切向量场，因而不存在截面的基.

同样，流形上余切向量，各种类型的张量的全体都构成向量丛，又如，设 M 为浸入于 Riemann 流形 N 的子流形，M 点的法向量的全体构成一个向量丛，称为 M 的法丛.

向量丛的转换函数 g_{pq} 一般取值于群 $GL(n)$，如果能够选取 $U_p \times F_p$ 的各截面基，使转换函数 g_{pq} 属于 $GL(n)$ 的某一子群 G，则称此向量丛可化约为群 G 的向量丛.

除向量丛之外，另一主要的纤维丛是关于某一 Lie 群 G 的主丛. 设 $M = \bigcup U_p$，定义乘积丛 $U_p \times G = \{(x, u_p) | x \in U_p, u_p \in G\}$. 若 $x \in U_p \cap U_q$，把 (x, u_p) 和 (x, u_q) 视为同一元素，但 u_p 和 u_q 之间由转换函数 $u_p = u_q g_{qp}(x)$ 所联系. 这样就能拼接成一个以 M 为底的群 G 的主丛 $P(M, G)$，自然，转换函数 $g_{pq}(x)$ 要满足

$$\begin{aligned} &g_{pq}(x) = g_{qp}^{-1}(x), \\ &g_{rq}(x) g_{qp}(x) = g_{rp}(x), \quad x \in U_p \cap U_q \cap U_r \end{aligned} \tag{43}$$

照样可定义投影算子 $\pi: P(MG) \to G$ 和截面 $\sigma: M \to P(M, G)$，二

者之间有关系 $\pi\circ\sigma = id_M$. 对于主丛而言，主丛化为乘积丛的充要条件是存在截面(当然应说关于 x 是连续的).

总之，纤维丛是以非常一般的形式，综合了微分几何和拓扑学中大量的主要研究对象，它的理论在整体微分几何中起了重要作用，后面我们还可见到，它还非常自然地用来表示有内在对称性的物理对象.

(2) 联络

联络的概念起源于 Riemann 几何学中的向量场的 Levi-Civita 平行移动和协变导数，不久就被扩充到联系于仿射变换群，共形变换群和射影变换群相应的仿射联络，射影联络和共形联络，而所有这些理论都可以纳入纤维丛上的联络的框架中.

设 $E \xrightarrow{\pi} M$ 是一向量丛，M 由一系坐标区域 U_p 所覆盖，E 为由乘积丛 $U_p \times F_p (p = 1,2,\cdots)$ 拼接而成. 设在 $U_p \times F_p$ 上已给好一组截面基 $\underset{p}{\sigma_\alpha}(x) (\alpha = 1,2,\cdots,m)$，则 $U_p \times F_p$ 上的线性联络由一组微分形式

$$\underset{p}{\Gamma} = \underset{p}{\Gamma_i} \mathrm{d}x^i \tag{44}$$

所组成，这里 $\underset{p}{\Gamma_i}$ 是取值于 $m \times m$ 阵的函数，$\underset{p}{\Gamma}$ 称为联络形式，方阵 $\underset{p}{\Gamma_i}$ 称为联络系数.

引进了联络，可以对截面进行求导的运算，设 $\underset{p}{\sigma}(x)$ 是一个截面，定义

$$D\underset{p}{\sigma}(x) = (\mathrm{d}v^\alpha + \Gamma^\alpha_{\beta i}(x) v^\beta \mathrm{d}x^i)\underset{p}{\sigma_\alpha} \tag{45}$$

称为 $\underset{p}{\sigma}(x)$ 的协变微分，而

$$\nabla_i \underset{p}{\sigma} = \left(\frac{\partial v^\alpha}{\partial x^i} + \Gamma^\alpha_{\beta i}(x) v^\beta\right)\underset{p}{\sigma_\alpha} \tag{46}$$

称为 $\underset{p}{\sigma}$ 的协变导数.

利用联络，还可以把 U_p 上的曲线 $x = x(t)$“水平提升”到 U_p

$\times F_p$ 上的一条曲线$(x = x(t), \underset{p}{v} = \underset{p}{v}(t) = v^a(t)\underset{p}{v_a}(t)$，这里的 $v^a(t)$ 是由微分方程

$$\frac{\mathrm{d}v^a(t)}{\mathrm{d}t} + \underset{p}{\Gamma^a_{\beta i}}(x(t))v^\beta(t)\frac{\mathrm{d}x^i}{\mathrm{d}t} = 0 \tag{47}$$

所决定，沿着提升后曲线的一系丛中向量，称为互相平行的.

利用联络，还可以引入曲率的概念，它的定义是

$$\underset{p}{\Omega} = \mathrm{d}\underset{p}{\Gamma} + \underset{p}{\Gamma} \wedge \underset{p}{\Gamma} = \frac{1}{2}\underset{p}{\Omega_{ij}}\mathrm{d}x^i \wedge \mathrm{d}x^j, \tag{48}$$

这里 Λ 是外积，d 是外微分.

$$\underset{p}{\Omega_{ij}} = \{\underset{p}{R^a_{\beta ij}}\} \tag{49}$$

对固定的 i, j 而言，它是 $m \times m$ 阵.

向量丛 $E \xrightarrow{\pi} M$ 上的联络是由各个 $U_p \times F_p$ 上的联络拼接而成，在拼接中对 $x \in U_p \cap U_q$，$\underset{p}{\Gamma}$ 和 $\underset{q}{\Gamma}$ 之间应满足

$$\underset{p}{\Gamma} = g_{pq}\underset{q}{\Gamma}g_{qp} + (\mathrm{d}g_{pq})g_{qp}, \tag{50}$$

这个关系保证了在 $U_p \cap U_q$ 中的曲线依 $\underset{p}{\Gamma}$ 和 $\underset{q}{\Gamma}$ 的提升的一致性. 曲率 $\underset{p}{\Omega}$ 的拼接关系是

$$\underset{p}{\Omega} = g_{pq}\underset{q}{\Omega}g_{qp}, \tag{51}$$

这样我们就有了向量丛 $E \xrightarrow{\pi} M$ 上的联络. 协变导数，水平提升，曲率等也都有了明确的定义.

向量丛 $E \xrightarrow{\pi} M$ 联络也可以有完全和坐标无关的描述. 记 $C^\infty(M,E)$ 为 E 的可微分截面的集合，$C^\infty(M,TM)$ 为 TM 的切向量场的集合，∇_X 为沿 X 方向的协变导数，那末对一切 $\varphi \in C^\infty(M, E)$，$X, Y \in C^\infty(M,TM)$，和 M 上可微函数 λ 成立：

$$\begin{aligned}
&(\text{I}) \quad \nabla_{X+Y}\varphi = \nabla_X\varphi + \nabla_Y\varphi. \\
&(\text{II}) \quad \nabla_X(\varphi + \psi) = \nabla_X\varphi + \nabla_X\psi. \\
&(\text{III}) \quad \nabla_{\lambda X}\varphi = \lambda\nabla_X\varphi. \\
&(\text{IV}) \quad \nabla_X(\lambda\varphi) = \lambda\nabla_X\varphi + (X\lambda)\varphi.
\end{aligned} \tag{52}$$

根据这 4 项性质，完全可以把联络确定下来.

主丛上的联络，也可以用拼接的方法定下来. 在每一乘积丛 $U_p \times G$ 上，用取值于群 G 的 Lie 代数 g 上一次微分形式

$$\underset{p}{b} = \underset{p}{b_i}(x)\mathrm{d}x^i \tag{53}$$

定义联络，以

$$\underset{p}{\Omega} = \mathrm{d}\,\underset{p}{b} + \frac{1}{2}[\underset{p}{b},\underset{p}{b}] = \mathrm{d}\,\underset{p}{b} + \frac{1}{2}[\underset{p}{b_i},\underset{p}{b_j}]\mathrm{d}x^i \wedge \mathrm{d}x^j \tag{54}$$

定义曲率，而拼接的关系式是

$$\underset{p}{b} = a\mathrm{d}(g_{pq})\,\underset{q}{b} - (\mathrm{d}g_{pq})g_{qp}. \tag{55}$$

利用纤维丛上的联络和曲率，可以对纤维丛的拓扑进行研究，利用陈省身和 A. Weil 的同态定理，可以通过适当选取的多项式来作出底流形 M 上的上同调类元素(称为示性类)，用以描述纤维丛的某些拓扑特性. 特别，设群 G 为 $U(1)$ 群，其 Lie 代数为一维的，以 $i = \sqrt{-1}$ 为其基，曲率形式就是 $\sqrt{-1}F_{ij}\mathrm{d}x^i\Lambda\mathrm{d}x^j$，第一陈示性类定义为 $\frac{1}{2\pi}F_{ij}\mathrm{d}x^i\Lambda\mathrm{d}x^j$. 若流形为紧致，其积分称为第一陈示性数.

设主丛 $P(M,G)$ 上已有一联络，M 为连通的流形，M 上取一定点 p，对始点和终点都是 p 的闭环路 l_p 作水平提升，每一 l_p 对应群 G 中一个元素 $u_{l_p} \in G$，这种对应称为和乐对应，u_{l_p} 全体所成集合构成一群，称为和乐群. 谷超豪证明，由和乐对应可以完全确定主丛的结构及其联络.

(3) 流形上的微分算子

流形上的一般的偏微分方程组可以用纤维丛概念来描述.

设 $E \xrightarrow{\pi} M$ 和 $K \xrightarrow{\tau} M$ 是两个以 M 为底的向量丛，用 ΓE 和 ΓK 记这两个丛的截面的集合. 在 M 的一个足够小的坐标邻域 U 上，E 和 K 的相应部分是乘积丛 $U \times F$ 和 $U \times H$，这里 F 和 H 是

m 维和 k 维的线性空间，在 $U\times F$,$U\times H$ 上各取一组截面基 $(\sigma_1(x),\cdots,\sigma_k(x))$ 和 $(\lambda_1(x),\cdots,\lambda_m(x))$，则 ΓE 和 ΓK 的元素可局部地表示为 $v^a(x)\sigma_a(x)$ 和 $w^a(x)\lambda_a(x)$. 设 L 是一个映照：$L:\Gamma E\to\Gamma K$，它能局部地表示为微分算子的形式

$$Lv=\sum_{\alpha_1+\cdots+\alpha_n\leqslant N}A_{\alpha_1\cdots\alpha_n}(x)\left(\frac{\partial}{\partial x^1}\right)^{\alpha_1}\cdots\left(\frac{\partial}{\partial x^n}\right)^{\alpha_n}v,\qquad(56)$$

那末就称 L 是一个微分算子，式中系数 A 是 m 列的 k 行的阵，N 是 L 的阶数，$\alpha_1,\cdots,\alpha_n$ 是非负整数. 如果 $m=k$，且 L 的主符

$$\sum_{\alpha_1+\cdots+\alpha_n=N}A_{\alpha_1\cdots\alpha_n}(x)\xi_1^{\alpha_1}\cdots\xi_n^{\alpha_n}\qquad(57)$$

当 $\xi_1,\cdots,\xi_n$ 不全为零时非退化，那末就称算子 L 是椭圆的. Riemann 流形上的 Laplace 算子就是椭圆算子的简单的例子，这时 E 和 K 都是乘积丛 $M\times R$.

算子 L 的核 $\ker L=\{\sigma\in\Gamma E\,|\,L\sigma=0\}$，即 L 的零空间，算子 L 的象空间是集合 $L\Gamma E=\{\lambda\in\Gamma K\,|\,\lambda=L\sigma,\sigma\in\Gamma E\}$. 当 M 是紧致无边的流形时，根据椭圆算子的性质，$\ker L$ 是有限维的，商空间 $\Gamma K/L\Gamma E$(称为 L 的上核，记为 $\operatorname{coker}L$) 也是有限维的. 算子 L 的指标是 $\ker L$ 和 $\operatorname{coker}L$ 的维数之差，即

$$\operatorname{Ind}L=\dim(\ker L)-\dim(\operatorname{coker}L).\qquad(58)$$

Atiyah 和 Singer 发现了著名的指标定理，证明了 $\operatorname{Ind}L$ 可以用丛 E、丛 K 和算子 L 的一些拓扑不变量表示出来. 这个定理的内涵非常广泛，把 Gauss-Bonnet 公式，Hirzebruch 符号定理和紧复流形的 Riemann-Roch 定理等作为其特殊情形，后来又在处理杨-Mills 方程的自对偶解的存在性和解的自由度等问题上发挥了重要的作用. 由于这个定理的证明十分复杂，寻求它的比较直接的、较为简化的证明仍然是很有意义的. 虞言林对 Dirac 算子和符号算子都进行了这样的工作，他和合作者还直接地证明 Lefschetz 公式.

微分几何中所出现的偏微分方程大多是非线性的，如极小曲

面，调和映照、杨-Mills 方程等，需根据方程的具体性质分别进行研究，这些方程都是拟线性的。此外完全非线性的方程，如 Monge-Ampère 方程，就更为复杂，需要用高度的分析技巧，近年来也已得到相当多的进展.

联系于椭圆算子，对抛物算子的研究也有所开展. Lorentz 流形和广义相对论中的许多问题，归结为双曲型方程的求解. 由于非线性项的影响，在许多情况下会出现奇性，从而整体解的存在性是一个重大的问题.

(4)杨-Mills 方程

在经典电磁理论中，电磁场强度用 4 维 Minkowski 时空中的二阶反称张量 f_{ij} 来表示，其中(f_{23}, f_{31}, f_{12}) 表示磁场强度，(f_{14}, f_{24}, f_{34}) 表示电场强度，由 Maxwell 方程可以知道，存在电磁势 $b_i(i = 1,2,3,4)$ 使得

$$f_{ij} = \frac{\partial}{\partial x_i}b_j - \frac{\partial}{\partial x_j}b_i, \tag{59}$$

而另一套 Maxwell 方程就是

$$\eta^{jk} \frac{\partial f_{ij}}{\partial x^k} = 0, \tag{60}$$

式中 $\eta^{jk} = 0(j \neq k), \eta^{11} = \eta^{22} = \eta^{33} = 1, \eta^{44} = -1$ 是 Minkowski 时空的度量张量. 此外，取 x 的任意函数 φ，作规范势的变换 $b_j \to b'_j$,

$$b'_j = b_j + \frac{\partial \varphi}{\partial x_j} = b_j + \frac{1}{\sqrt{-1}}\left(\frac{\partial}{\partial x_j}e^{i\varphi}\right)e^{-i\varphi}, \tag{61}$$

电磁场并没有发生变化. 这种形式的变换称为规范变换. $e^{i\varphi}$ 是群 $U(1)$ 的元素，用物理学的术语来说，光子拥有群 $U(1)$ 的内禀自由度.

杨振宁和 R. S. Mills 从中子和质子在某种程度上的群 $SU(2)$ 对称性，预见到有和 $SU(2)$ 相应的场，它的势函数是取值于 Lie 代

数 $SU(2)$ 的函数 b_i，场的强度由取值于 $SU(2)$ 的

$$f_{ij} = \frac{\partial b_j}{\partial x^i} - \frac{\partial b_i}{\partial x^j} + [b_i, b_j] \tag{62}$$

所成，他们提出了场的作用量(今称杨-Mills 泛函)，

$$YM(b) = \frac{1}{4}\int_M tr(f_{ij}f^{ij})\mathrm{d}x^4),$$

场方程就是这作用量的 Euler 方程：

$$\eta^{jk}\left(\frac{\partial f_{ij}}{\partial x_k} + [b_k, f_{ij}]\right) = 0. \tag{63}$$

这情况和电磁场十分类似，但由于 $SU(2)$ 是不可换的，所以在场的强度和场方程中，都出现了换位运算[,]的项. 此外，规范变换具形式

$$b_j' = (adw)b_j + (\partial_j w)w^{-1}, \tag{64}$$

这里 w 是取值于 $SU(2)$ 的函数. 这理论就称为 $SU(2)$ 群的规范场理论.

这个理论体系可以推广到任何 Lie 群，从 60 年代以来它在建立弱相互作用和电磁相互作用的理论，强子的内在结构，都取得很大的成功，且已成为研讨四种基本相互作用(引力，弱，电磁，强)的统一理论的重要工具.

70 年代中，数学中纤维丛理论和规范场的关系已被沟通，明确了规范场就是联络，场强就是曲率，杨振宁还指出了非乘积丛在物理学中的作用，例如第一陈示性数就是磁荷. 与此同时，许多数学家也注意到了物理学中的规范场理论，过去数学家在纤维丛的研究中，从未想到要引入一个决定联络的偏微分方程，而这样的方程却被物理学家杨振宁与Mills成功地找到了. 数学家们着重讨论 4 维欧氏空间和 Riemann 流形上杨-Mills 方程，得出了丰硕的成果.

人们首先感兴趣的是欧氏空间 R^4 上杨-Mills 方程的有限能量解. 利用球极投影和可去除奇点定理，可把问题化到 S^4 上求解，

于是有关紧致流形上的椭圆算子 Atiyah 和 Singer 的理论就起了重要的作用，特别是对这种方程的一大类具有特殊性质的解(称为自对偶和反自对偶解，也称为瞬子解)，非但定出了解的自由度，而且利用复几何和代数几何的技巧，给出了它们的有效的构作方法. 这类解的重要意义还在于，使杨-Mills 泛函达到局部最小的解都是瞬子解.

随后又讨论了比较一般的 4 维 Riemann 流形上的杨-Mills 方程的瞬子解的存在性问题，解所成的拓扑空间(称为模空间)的结构也有很大的进展. 最使人惊异的是，这种方程在研究 4 维流形的微分拓扑中起了重要作用，英国的 K. Donaldson 利用它来研究 4 维流形的微分拓扑，美国的 M. Freedman 进而发现 R^4 的非通常的微分结构(对任何 R^n，$n\neq 4$，只有通常的微分结构)，他们二人都得到了 1986 年的 Fields 奖.

在规范场的数学结构方面，我国学者谷超豪等和杨振宁教授从 70 年代中期开始进行了合作研究，取得了系统的成果. 其中包括：规范场的环路位相因子方法的建立，场强对规范场决定作用的分析，球对称规范场的决定和 Cauchy 问题解的存在性以及有限能量和慢发散能量静态解的不存在性，等等. 其后，葛墨林研究了自对偶解的潜在对称性和无限维代数结构，胡和生等给出了在弯曲时空中杨-Mills 场有质量规范场及手征场的若干不存在性定理，沈纯理将杨-Mills 理论和两个复流形间的全纯映照联系起来，计算了某些模空间，讨论了它们的拓扑性质.

后　　记

这篇稿子是三年以前写的. 这几年来微分几何发展非常迅速，文中所涉及的各个方面，都有相当大的进展. 现在只能作一些非常简略的补充说明.

整体微分几何的发展有好几个方面的推动力. 第一,现代分析正和微分几何紧密结合,许多分析学家关心几何中的问题,利用分析工具,得出许多深刻的成果. 几何分析(geometrical analysis) 这一术语已广泛流行,被视为现代数学发展的主流方向之一. 第二,理论物理继续需要运用微分几何作为数学工具. 规范场理论的成功已是众所周知. 有些探索性理论(如超弦),也用到了很深刻的数学,虽然它们的实验根据尚嫌不足,在物理学界尚有不同看法,但对数学发展的推动作用却十分强烈. 此外,丘成桐收集和发表的一系列的微分几何待解决的问题,吸引了许多学者,其中许多问题,都得到很好的进展.

具体的成果非常多,方面也广,这里只能比较随机地列举一些带有代表性的成果. 在国际上(一) 在 Riemann 流形 M(n维) 到欧氏空间 R^N 的等距嵌入的研究中,已经证明如 $N = \max\{n(n+5)/2, n(n+3)/2+5\}$,这种嵌入一定能实现. (二) 在 Riemann 流形曲率和拓扑的关系方面,人们充分注意到曲率和其他几何量,如流形的体积、半径等,对流形的拓扑都能有相当的影响,成果很丰富. 举一个例子:Grove 等人得到如下的大体积定理. 设 M 为 n 维 Riemann 流形($n \geqslant 2$),k 为一实数,R 为一正数(当 $k > 0$ 时,要求 $R \leqslant \frac{\pi}{2}\sqrt{k}$),必存在一实数 $\varepsilon = \varepsilon(n,k,R) > 0$,使得:当 riem $M \geqslant k$,M 的半径 $\leqslant R$,M 的体积 $\geqslant V_k^n(R) - \varepsilon$ 成立时,M 的拓扑必为 S^n 或 RP^n,这里 V_k^n 是曲率为 k 的 n 维空间形式中以 R 为半径的测地球的体积. (三) 发现了液晶分子的排列和调和映照有密切关系,从而对调和映照特别是边值问题奇点集的结构有很多推进,同时调和映照流也得到了许多研究,如弱解的存在性,正则性,破裂或渐近形态等. (四) 对接近于平坦的初始条件,已证明 Einstein 方程的整体解的存在性,也即证明了 Minkowski 时空的稳定性. 又已证明 Yang-mills 方程无自对偶性的整体解的存在性. 特别是丘

成桐等得出了 Einstein 方程和 Yang-mills 方程的偶合方程无奇性静态解存在，在物理上很具吸引力.

由于天元基金和国家自然科学基金的支持，由于国际交流的开展，在过去几年中国内的研究有许多可喜的、引起国内外同行关注的成果. 在调和映照方面，忻元龙发展了等变映照的概念，证明了高维圆盘到高维球面的轴对称调和映照的“大边值”问题解的存在性. 他又证明了 $\pi_{2m+1}(S^{2m+1})$ 中任一奇数类一定存在调和代表元. 在子流形方面，沈一兵等得到高维球面、复射影空间中某些极小子流形的 Ricci 曲率和数量曲率的目前最佳的 pinching 常数，他又给出了复射影空间中 2 型极小曲面的完全分类. 彭家贵等给出了 Bonnet 曲面和三次极小锥的完全分类，肖良给出了球面中等参子流形多项式表示和复射影空间等参子流形的分类. 在规范场方面，沈纯理和陈韵梅合作研究了 Yang-Mills 流，讨论了小范围时间解的存在性和演化中的正则性和奇点集. 胡和生研究 Yang-Mills 方程的团块现象，她证明在能量无限但能量发散较缓时也不能有团块现象，她还得一些团块解的实例. 胡和生把孤立子理论运用于微分几何，用 Darboux 变换显式地构造了射影微分几何中苏链(又称 Фиников 构图)，并证明 $2n(n>1)$ 为周期的序列的存在. 谷超豪与胡和生合作，对手征场($R^{1+1}\to SU_2$)得出单孤立子和多孤立子解的显式表达式，并证明这种孤立子的相互作用是弹性散射. 他们还合作得出推广的非线性波动方程和 Sine-Gordon 方程求解的显式方法等. 虞言林研究了 Signature 算子局部指标定理的非直接证明，指出前人证明中的疏漏. 在仿射微分几何中，李安民作了一系列的工作，特别是对于仿射球的分类问题，他发现和补正了前人的工作的疏漏，从而完成了对 Calabi 所作的有关猜想的证明. 李安民还证明了：仿射完备极大曲面如果其仿射法线不取 4 个一般位置的方向，则一定是椭圆抛物面.

更值得指出的是，国内有若干从事非线性分析和偏微分方程

研究的专家，也正在关心和研究微分几何方面的问题，得出了多项重要成果. 张恭庆、丁伟岳等对变分问题和半线性椭圆型方程的研究，在众所关注的多项几何问题中得到应用，取得重要结果. 王光寅、吉敏关于黎曼流形中极小曲面存在性的结果，引起了广泛的注意与重视. 洪家兴利用非线性混合型方程研究曲率变号的二维流形的嵌入问题，取得深刻的结果.

在这期间，不少青年学者也进入这个领域，有可喜的成果. 这一切都表征着这门学科在我国已有重要的进展和蓬勃的生命力.

参 考 文 献

1. 华罗庚，苏步青主编，中国大百科全书·数学，1988.
2. Kobayashi, S. and Normizu, K., Foundations of Differential Geometry, 1969, Interscience Publishers.
3. 日本数学会编，数学百科全书(中译本)，1984，科学出版社.
4. 陈省身，陈维桓，微分几何讲义，1983，北京大学出版社.
5. 丘成桐，孙理察(R. Schoen)，微分几何，1988，科学出版社.
6. 苏步青，胡和生等，微分几何学，1979，人民教育出版社.
7. 谷超豪，On Classical Yang-Mills Fields, Physics Reports, 1981.

周 青
（华东师范大学数学系）
Zhou Qing
(Department of Mathematics,
East China Normal University)

什么是低维拓扑学
What Is Low-dimensional Topology

中国数学发展的若干主攻方向
Some Main Aspects on Development
of Mathematics in China

§1 什么是低维拓扑学?

拓扑学是20世纪理论数学中发展最为迅速的分支之一. 60年代是代数拓扑学黄金时代. 这样讲不仅是由于这些理论本身的完美,也由于它们已经渗透到数学的许多别的分支中去了. 代数拓扑学的工具对于研究维数大于或等于5的流形结构起了本质的作用. 许多高难的问题,由于代数拓扑学的发展而变得明朗了. 但是对于了解我们所生活的维数的空间,这些工具没有给我们太多的帮助. 所以一些数学家们开始问自己:拓扑学是否走得太远了?同时他们也逐渐把自己的兴趣转向2维、3维、4维流形的研究. 所谓低维拓扑学,并没有一个很明确的定义,通常的理解就是2维、3维和4维流形的拓扑学研究,它是近年来数学中最活跃的分支之一.

§2 早期的低维拓扑学

上述意义的低维拓扑学并不是20世纪60年代才开始的,它可以追溯到19世纪的闭曲面的分类. 那时候的低维拓扑学的研究主要关心2维的理论. 直到20世纪50年代,数学家们对于3维流形知道的还少得可怜.

有一类3维流形,叫做Seifert流形,算是了解得比较清楚了. 粗略地说,Seifert流形是这样的一种流形,它们可以看作是一组闭曲线的并,而且这些闭曲线在局部是很规则地排列起来的. 例如在一个实心环上,我们可以给出下面的Seifert纤维化. 一个实心环可以看作是一簇平行的环面加上实心环的中心线组成的. 图1中我们画出了其中一个环面上的两种不同的纤维化. 这两种Seifert纤维化的不同之处是,图1*b*)中的纤维化不是一个乘积纤维化,中心线是一个奇异纤维,因为这个纤维绕实心环的洞是一圈,而它旁

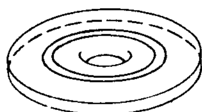


图 1a)



图 1b)

边的别的纤维都绕许多圈. 如果我们把每一个闭曲线看作一个点，那末这个 Seifert 流形的商则是一个 2 维曲面，换句话说，Seifert 流形是一个用曲面参数化的闭曲线族，用关于这个曲面和那些奇异纤维的信息可以完全刻画 Seifert 流形.

另一个比较早就知道的有趣的定理叫做连通和分解定理. 我们先看看类似的 2 维情形. 如图 2. 在一个 2 维曲面 F 上，如果有一

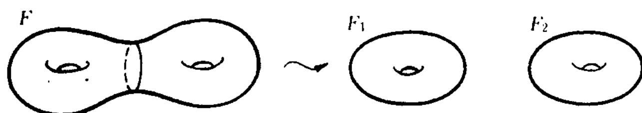


图 2

条分离的简单闭曲线 C(分离指的是沿这条曲线割开 F，我们将得到两块)，沿这条曲线 C 割开 F，我们得到两个带圆周边界的曲面，把这两个圆周边界分别捏成点，我们得到两个曲面 F_1，F_2. 这个过程叫做 F 的一个连通和分解. 如果我们用 3 维流形中的分离球面，代替 2 维流形中的分离简单闭曲线，我们就得到 3 维流形的连通和分解. 如果我们允许得到的流形中可以有新的 3 维球面出现，那末这个分解可以无休止地进行下去. 因为我们总可以在 3 维流形中取一个小的实心球，沿其边界球面割开再把边界捏成一点，这小实心球就形成一个新的 3 维球面. 有新的 3 维球面出现的连通和分解叫做平凡的分解，连通和分解定理说，对每一个 3 维流形，我

们可以做若干次非平凡的连通和分解，使得最后得到的那些流形不再有非平凡的分解，而且本质上这种分解是唯一的. 这个定理类似于整数的素因数分解定理. 所以我们又称那些不再有非平凡的连通和分解的3维流形为素的3维流形. 根据这个定理，了解3维流形，我们只须要了解素的3维流形就足够了.

§3 60年代和70年代的组合3维拓扑学

60年代中期，Haken引入了一个很重要的概念——“足够大的3维流形”. 这种流形中，有一张很好的嵌入的不平行于边界的曲面F(假定它不是球面)，而且F上的每条可以在这个3维流形中收缩成一点的简单闭曲线，在F中也可以收缩成一点. 具有这种性质的曲面叫做一张不可压缩的曲面. 如果这张曲面是双侧的，那么这张曲面的基本群就嵌入到3维流形的基本群成为子群. 这样，这种流形的基本群足够大，包含了一个曲面群，这也是这个名称的由来. 通常称足够大的素的3维流形为Haken流形(可以证明：每一个带边界的素的3维流形都是Haken流形)，除非它是一个3维实心球. 对于Haken流形，我们有一个很好的分类定理：同伦等价的Haken流形是同胚的.

证明这个分类定理，最重要的工具是hierarchy. 用2维作对比，我们可以沿一些闭曲线和弧割开一个曲面，使它成为一个多边形. 在3维中，我们要沿一些曲面去割3维流形. Haken流形中存在不可压缩的曲面，而且可以证明这些曲面是由它们的基本群在3维流形基本群中的象决定的. 沿一张不可压缩曲面割开原来的流形后，我们仍然得到一个Haken流形，或是3维实心球，因为这时候流形带边了，素的条件可以很容易地从这张割开曲面的不可压缩性推出来. 我们可以继续这样地割开3维流形，除非我们得到一组3维实心球. 关键的一步是证明我们不能无休止地这样做下去，

所以有限步以后，我们得到一组 3 维实心球. 这个一步步割开 3 维流形的过程叫做一个 hierarchy. 现在我们来看一下分类定理的证明，假定 M,N 是两个同伦等价的 3 维 Haken 流形. $f:M\to N$ 是一个同伦等价、固定 N 上的一个 hierarchy，第一步是沿一张不可压缩的曲面 S_1 割开 N，得到的流形记做 N_1. 注意 f_* 在基本群上是同构，所以 $f_*^{-1}(\pi_1(S_1))\subset\pi_1(M)$ 是一个曲面群，我们可以找到一张不可压缩的曲面 $F_1\subset M$，使得 $\pi_1(F_1)=f_*^{-1}(\pi_1(s_1))$，然后同伦 f 到 f_1，使 $f_1:F_1\to S_1$ 是一个同胚，而且 f_1 诱导出 $M-F_1\to N-S_1$ 上的一个同伦等价，归纳地做下去，我们可以得到一个在两组实心球上的同伦等价，而且在边界上是同胚，所以这个同伦等价可以很容易地用同胚去代替. 把这些同胚粘起来，我们就得到一个 M 到 N 上的同胚.

这个阶段中，3 维流形的组合研究有了很大的发展，纯粹组合方法的最后一个漂亮的定理叫做环面分解定理，想法有点像连通和分解，不过现在我们是沿一些不可压缩的环面去割一个素的 3 维流形，割开以后我们得到一些 Seifert 流形和其他一些不再包含任何不平行于边界的不可压缩环面的流形. 有意思的是这种分解还是在本质上唯一的. 前面我们已经说过，Seifert 流形算是了解得很好了，余下的是了解那些 3 维流形，它们不包含任何不平行于边界的不可压缩的环面.

§4 Thurston 对曲面的研究工作

当然，在 70 年代以前的 3 维流形的研究方法也不仅仅是组合方法，例如 Novikov 的叶结构理论，但主要是组合方法.

研究方法上的真正的发展是 70 年代中期开始的，而且最初的那些理论都是和 Thurston 的名字连在一起的.

Thurston 早期的工作是关于流形上的叶结构的存在性的. 70

年代中期，他利用一些分析学家们熟悉的概念，例如 Teichmuller 空间，拟共形映射等，来研究 2 维曲面上映射的分类，发展了一整套理论，许多搞 Riemann 曲面的数学家们为之一振，使拟共形映射的研究得到了新的活力.

我们来看看 Thurston 的主要的方法是什么. 首先我们描述一下一个亏格大于 1 的闭曲面 Teichmuller 空间. 假定 F 是一个这样的曲面，粗略地讲，F 的 Teichmuller 空间 $\mathscr{T}(F)$ 是 F 上的双曲结构的同痕类的全体. 它可以用下面的方法构造出来：如图 3，如果 F 的亏格是 g，我们可以找到 $3g-3$ 个简单闭曲线，沿这些曲线割开

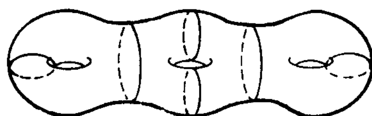


图 3

F，每一块是一个带两个洞的圆盘. 在一个带两个洞的圆盘上的双曲结构可以由它的三个测地边界的长度决定. 当我们把这些双曲块粘合起来时，我们还需要有一个角度去度量这样的粘法. 由于我们考虑的是双曲结构的同痕类，所以图 4 中的两种粘法被认为是

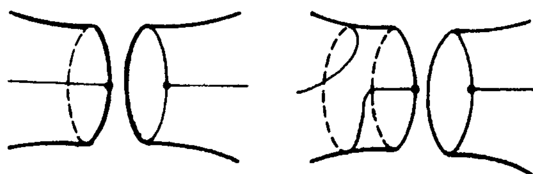


图 4

不一样的，这样一来每条简单闭曲线提供了两个参数，一个是长度，另一个是粘合的转动角度，或者说沿 $\mathscr{T}(F)$ 是同胚于 R^{6g-6}

的. F 上的每一个自同胚诱导 $\mathscr{T}(F)$ 上的一个自同胚，而且同痕的自同胚给出 $\mathscr{T}(F)$ 上相同的映射，这个曲面上的自同胚的同痕类构成一个群，通常叫做映射类群.

Teichmuller 空间的结构是很早就清楚的. Thurston 的贡献在于他首先给出了 $\mathscr{T}(F)$ 的自然的紧致化，使得映射类群在 $\mathscr{T}(F)$ 上的作用可以延拓到边界上去. 他是这样考虑的. $\varphi(F)$ 表示 F 上的非平凡的简单闭曲线的同痕类的集合，对于 $\varphi(F)$ 中的两个元素 α,β，定义 $i(\alpha,\beta)$ 是在 α,β 的同痕类中简单闭曲线的的最少相交数. Thurston 证明了，对每个 $\alpha \in \varphi(F)$，我们可以找到一个 $\beta \in \varphi(F)$，使得 $i(\alpha,\beta) \neq 0$，而且对于两个不同的元素 $\alpha_1,\alpha_2 \in \varphi(F)$，我们可以有一个 $\beta \in \varphi(F)$，使得 $i(\alpha_1,\beta) \neq i(\alpha_2,\beta)$. 这样一来，$i$ 便定义了一个单的映射 $i_*:\varphi(F) \to \mathscr{P}R^{\varphi(F)}$. 而且象的闭包 $P\varphi(F)$ 是紧致的，这就是 $\mathscr{T}(F)$ 的一个自然的边界. 为了能看清楚这一点，我们可以把 $\mathscr{T}(F)$ 也嵌入到 $\mathscr{P}R^{\varphi(F)}$ 中去. 对每一个 $\alpha \in \varphi(F)$，$\gamma \in \mathscr{T}(F)$，$i(\gamma,\alpha)$ 表示 α 中的简单闭曲线在 γ 这个几何结构下的长度的下确界，这个 i 有着与前面的 i 几乎相同的性质，同样也给出一个到 $\mathscr{P}R^{\varphi(F)}$ 中的嵌入，象的闭包则是象与 $P\varphi(F)$ 的并. 事实上，我们还可以证明，这个紧致化同胚于一个 $6g-6$ 维的实心球.

这个紧致化的一个直接推论是：F 上的任何一个自同胚，或者在 Teichmuller 空间上有不动点，或者在 $P\varphi(F)$ 中有不动点. 在第一种情况，Teichmuller 空间中的不动点给出 F 上的一个双曲结构. 而且这个自同胚同痕于一个等距，第二种情况要分析那些在 $P\varphi(F)$ 中的不动点. 可以证明这时候自同胚根据有一个不动点或两个不动点分成是可约的，或者是伪 Anosov 的. 这就是 Thurston 的关于曲面上自同胚的分类定理.

§5 3维流形上的几何结构

Thurston 是一个神秘的人物，他不屑于写他自己做的工作，于是很多人就帮他写，正当人们热衷于写他在曲面上的工作时，他自己却悄悄地开始了对3维流形的几何结构的研究.

他证明了如果一个 Haken 流形 M 不包含任何不平行于边界的环面，那么 M 的内部允许一个完备的双曲度量，除非 M 是一个 Seifert 流形或者是一个圆周上的环面丛. 证明还是在 hierarchy 上用归纳法. 我们来看归纳法中的一个情况，取 M 中的一个不可压缩曲面 F，然后沿 F 割开 M，我们能得到两个允许双曲结构的3维流形，想法是把这两个双曲几何粘起来，得到 M 上的一个双曲结构. 最简单的情形是我们可以找到两个全测地嵌入的曲面平行于那些粘合的边界，而且粘合映射是一个等距，即我们可以直接把这两个双曲结构粘在一起. 一般说来，我们不能保证粘合映射是等距，甚至不是永远可以找到这样的全测地嵌入曲面的. 所以我们要了解怎样去改变这些几何结构，使得它们足够好. 可以证明一个双曲流形上的几何结构1-1对应于它们在非环面边界上诱导的复结构. 所以双曲模空间是一些 Teichmuller 空间的乘积. Thurston 用不动点定理证明了，我们总可以形变这两个双曲结构，使得在无穷远的粘合面上的复结构是一致的，而且粘合映射是一个共形映射. 这样 Maskit 的定理告诉我们，可以用这两个双曲结构得到 M 上的一个双曲结构.

尽管 Haken 流形算是有了一个分类定理了，不过双曲结构的存在性也还是很重要的. 有了这个结果，3维的拓扑学大大地向前迈了一步. 对于 Seifert 流形，我们有6种几何学，而且圆周上的环面丛又有一种几何，3维的情况我们一共有8种几何. Thurston 根据他的定理，提出了一个猜测：任何3维流形，经过连通和分解和

环面分解后，得到的块上都有几何结构. 这个猜测是 3 维拓扑学中最主要的猜测了，过去的很多著名猜测，包括至今悬而未决的三维的 Poincaré 猜测，都是 Thurston 猜测的特别情形.

Thurston 还证明了一个关于 3 维流形上的几何结构的结果，如果一个素的 3 维流形允许一个有限群的作用，而且这个作用有大于 1 维的不动点集，那么这个 3 维流形有上面讲的那种几何分解，而且这个群作用限制在分解后得到的每个块上是等距作用. 那一段时间是 3 维流形的很光辉的时代.

§6　极小曲面的应用

70 年代末和 80 年代初时的另一个热点是极小曲面理论在 3 维拓扑学中的应用. 这套方法首先是由丘成桐和 Meeks 开始的. 当然，证明极小曲面的存在性也是很困难的，不过那不是低维拓扑学所关心的. 他们考虑一个具有最小面积的非平凡的球面，看上去这种曲面应该有最少的自身相交的性质. 事实上，他们证明了，如果这个球面同伦于一个嵌入的球面，那么这个最小面积的球面本身也是一个嵌入或者是到一个嵌入的射影平面上的 2 叶覆盖. 更有意思的是两个这样的球面或者重合或者相互不交. 这样一来，我们可以用这个结果到有限群在 3 维流形上的作用. 如果我们有一个有限群 G 作用在 3 维流形 M 上，不难在 M 上给出一个度量，使得 G 是一个等距群. 利用极小曲面，我们可以找到一组等变的球面. 这样，如果 M 有非平凡的连通和分解，那么我们可以找出一个等变的连通和分解. 这叫做等变的球面定理.

极小曲面的理论在等变理论中有许多应用，主要的哲学根据是极小曲面应该有最少相交的性质，也有人研究等变的大亏格的极小曲面. 不过在这些研究中，等变的球面定理是最有意义的.

§7 单连通闭 4 维流形的分类

前面我们主要讨论 2 维和 3 维流形的研究. 4 维流形的早期研究的确很少. 假定 M 是一个闭的 4 维流形，那么在第 2 个上同调群的自由部分 $FrH^2(M, Z)$ 上的上积：

$$I_M: FrH^2(M, Z) \times FrH^2(M, Z) \to H^4(M, Z) = Z$$

给出一个整二次型，这个整二次型叫做 M 的 2 次相交形式.

Poincaré 对偶定理说这个相交形式是么模的，整二次型当然可以看作是一个实的二次型，所以符号差 $\sigma(I_M)$ 是一个流形的不变量，通常记作 $\sigma(M)$. 整二次型也有一些不同于实二次型的地方，如果对于任何 x，$I_M(x,x) \equiv 0(\mathrm{mod}2)$，我们称这个整二次型是偶的. 一个二次型是偶的，当且仅当它的矩阵表示的对角线上的元素全是偶数. 例如

$$H = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad E_8 = \begin{bmatrix} 2 & 1 & & & & & & \\ 1 & 2 & 1 & & & & & \\ & 1 & 2 & 1 & & & & \\ & & 1 & 2 & 1 & & & \\ & & & 1 & 2 & 1 & 0 & 1 \\ & & & & 1 & 2 & 1 & 0 \\ & & & & 0 & 1 & 2 & 0 \\ & & & & 1 & 0 & 0 & 2 \end{bmatrix}$$

是偶形式. 不是偶的二次型叫做奇二次型，奇形式和偶形式在整数环上是永远不会等价的.

人们很早就认识到这个二次相交形式是单连通闭 4 维光滑流形的一个重要不变量，它决定了这种流形的伦型. 事实上，它决定了这些流形的 h- 协边类. 过去人们总认为 4 维是一个最黑暗的维数，距离得到一个好的分类定理太远. 1980 年，Freedman 得到一个

单连通闭 4 维流形的一个完整的拓扑分类定理. 一下子,这个最黑暗的维数变得光明了. 他的定理说:从闭 4 维单连通流形 M 到它的相交形式 I_M 的映射是满的,而且是 2 到 1 或 1 到 1 的,依赖于 I_M 是奇形式或偶形式.

这个定理的证明与高维的类似结果的证明很像. 高维的方法不能用到低维上来,主要是因为 Whitney 技巧在低维时失效了. 1973 年,Casson 有一个方法,可以说是一个同伦范畴中的 Whitney 技巧. Casson 当时就意识到,如果这也是一个光滑范畴中的 Whitney 技巧的话,那么 4 维可以和高维做的一样好. Freedman 花了很长时间研究 Casson 的这个方法,最后他证明了,Casson 的方法事实上是一个拓扑范畴的 Whitney 技巧,这个证明用了一种比较几何的点集拓扑学 ——Bing 的拓扑学. Whitney 技巧在拓扑的范畴中可以通过,那么 Freedman 的定理就得到了.

§8 4 维光滑流形拓扑

尽管 Freedman 的定理给出了所有的 4 维闭的单连通拓扑流形,我们仍然不清楚有多少光滑流形,也不清楚拓扑流形与光滑流形的差别. Rohlin 有一个定理说:单连通的光滑闭 4 维流形 M,如果 I_M 是一个偶形式,那么 $\sigma(M) \equiv 0 (\mathrm{mod}\ 16)$. 初等的整二次型定理说,偶形式的符号差模 8 为 0. 这样我们知道在具有偶形式的闭单连通 4 维流形中,至少有一半是不可光滑化的. 事实上,在那些具有奇形式的流形中,也至少有一半是不可光滑的. Freedman 的 2 对 1 的映射可以表述成如下的形式,如果我们乘上一个 S^1 得到的一个是可光滑化的 5 维流形,另一个是不可光滑化的,是否还有别的不可光滑的 4 维拓扑流形?

1982 年,Donaldson 在他做研究生的第二年,得到了另一个光滑化的阻碍,他证明了:一个光滑的单连通的闭 4 维流形,如果具

有正定的相交形式，那么这个相交形式一定是平方和. 我们前面提到的偶形式 E_8 是一个正定的偶形式，但它不能等价于一个平方和. 因为平方和是一个奇形式. Rohlin 定理说 E_8 不能是一个单连通闭 4 维流形的相交形式，而 Donaldson 定理说 $E_8 \oplus E_8$ 也不能成为一个单连通闭 4 维流形的相交形式. 令人吃惊的不仅仅是这个定理本身和它的很多重要的推论，而且方法上也是很特别的，他把两个原来人们认为不相干的分支，拓扑学和规范场理论，联系在一起了.

固定一个在 M 上的复 2 维向量丛 E，它的第二个 Chern 类是 1. 考虑全体自对偶的 Yang-Mills 连络在 $SU(2)$ 规范变换下的等价类所成的空间 $\mathscr{M}$，Atiyah-Singer 指标定理说，这个 $\mathscr{M}$ 是一个带有限个奇点 $p_1, \cdots, p_s$ 的 5 维光滑流形，每个奇点对应于一个可约的连络. 拓扑上，一个可约的连络对应于这样的 2 维同调类 $\pm u \in H^2(M,Z)$，$I_M(u,u) = 1$，所以奇点个数 $s = \frac{1}{2} \# \{u \in H^2(M,Z), I(u,u) = 1\}$. 在奇点附近，$\mathscr{M}$ 是一个 CP^2 上的锥.

我们还需要了解 $\mathscr{M}$ 在无穷远处的情况. 这里最基本的工作是 Uhlenbeck 做的. 她证明了如下的正则性定理：如果一个序列 $\nabla^i \in \mathscr{M}$ 趋向无穷远处，那么我们找到一个子序列 ∇^{i_j}，除了在 M 上的一个点 P 外，∇^{i_j} 收敛到一个 M 上的平坦连络，而且 ∇^{i_j} 的能量逐渐集中到 P 点的一个越来越小的邻域内. M 上的平坦连络是 1-1 对应于 $\pi_1 M$ 在 $SU(2)$ 中的表示的，因为 M 是单连通的，我们只有一个平坦的连络，所以能量集中到 P 点的越来越小的邻域中的 Yang-Mills连络可以用实数来参数化，Taubes给出了这样的构造. 对M上的每一点P，与一个充分小的正数，存在一个这样的能量集中在P点的ϵ球中的自对偶的Yang-Mills连络. Donaldson利用这个构造，证明了在无穷远处，$\mathscr{M}$ 同胚于 $M \times (0,i)$. 这样我们可以在$\mathscr{M}$中找到一个紧的5维流形$\mathscr{M}$，使得$\partial \mathscr{M}_0 = M + pCP^2 + q$

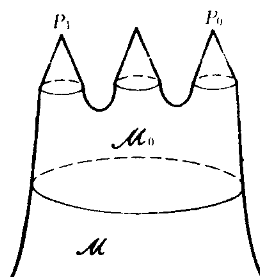


图 5

$\cdot(-CP^2)$，其中 $p+q=s$. 我们知道 $\sigma(\partial\mathscr{M}_0)=0$，所以 $\mathrm{rank}H^2(M,Z)=\sigma(M)=q-p\leqslant q+p=s\leqslant\mathrm{rank}H^2(M,Z)$ 这推出 $s=\mathrm{rank}H^2(M,Z)$. 一个初等的代数引理说，

$$\frac{1}{2}\#\{u\in H^2(M,Z);I(u,u)=1\}=\mathrm{rank}H^2(M,Z),$$

当且仅当 $I(\cdot,\cdot)$ 是一个平方和.

这个定理有一个令人惊奇的推论：R^4上存在至少两个不同的光滑结构. 这更显出 4 维是个独特的维数，因为人们早就知道，其他维数的欧氏空间上都只有唯一的光滑结构. 现在人们已经知道 R^4上不同的光滑结构多极了，至少要用两个实数来参数化.

Donaldson 是一位大师，他用规范场理论，还给出了 4 维拓扑学中的很多别的结果. 其中有一个是在 $CP^2\#q(-CP^2)$ 上有两个不微分同胚的复结构，后来被别的数学家推广为存在无穷多个彼此不微分同胚的复结构. 4 维流形上的光滑结构的理论实在是太丰富了.

§9 纽结的 Jones 多项式和 Witten 的工作

低维拓扑学还有一个重要的对象，就是所谓的纽结. 纽结是一

个圆周在3维球面S^3中的拓扑象，而若干个圆周的拓扑象则称为一个链环. 这些拓扑象本身就是一个或一组圆周，没有什么拓扑学可言. 可是研究它们在S^3中的放置方法却大有学问. 图6是最基本



图 6

的两个不同的纽结，从经验可知它们是不同的，但是如何从数学上来区别它们呢？最早研究的是它们的补空间，事实上，最近Gordon证明了纽结由它们的补空间一意决定. 补空间的基本群就是纽结的一个不变量. 而一般说来，这个群是非交换的. 判定两个非交换群是否同构，是相当困难的，所以这是一个不太好用的不变量. Alexander曾经得到过一个多项式不变量. 尽管这个多项式没有补空间的基本群所包含的信息多，但是比较两个多项式是否相同太方便了. 80年代初期，Jones很偶然地发现了一个纽结的多项式不变量，现在被称为纽结的Jones多项式. Jones是一位分析学家，当时他的兴趣是有限Hecke代数的表示，一次在一个讨论班上，他在黑板上写下一组关系，听众中有人指出，这组关系很像辫群的关系. 辫群是一个与纽结关系很密切的东西，此后他潜心研究，终于发现了纽结Jones多项式. 今天这个多项式已经被拓扑学家弄得十分简单，当时的确是拓扑学上的一件大事，人们没有想到从算子代数的角度可以给出一个纽结的新的不变量. 无独有偶，人们一样没有想到，1989年一个理论物理学家Witten用3维的规范场理论给出了Jones多项式的一个解释. 设K是S^3中的一个纽结，在S^3上取一个平凡的$SU(N)$主丛E，而且在E上放一个连络A. 我们可以沿着纽结K计算A的holonomy，这样得到$SU(N)$中的一个元素，这个元素的共轭类由纽结一意决定，这个元素的迹叫做

Wilson 线

$$W(K)=Tr\mathbf{P}\exp\int_K A_i dx^i.$$

然后，我们计算 Feynman 道路积分

$$\int D\mathscr{A}\exp(i\mathscr{L})W(K).$$

这里 $D\mathscr{A}$ 表示在全体连络的规范轨道上求的 Feynman 积分，而 Lagrange 作用量取的是 Chern-Simons 形式

$$\mathscr{L}=\frac{k}{4\pi}\int_M Tr(A\wedge dA+\frac{2}{3}A\wedge A\wedge A).$$

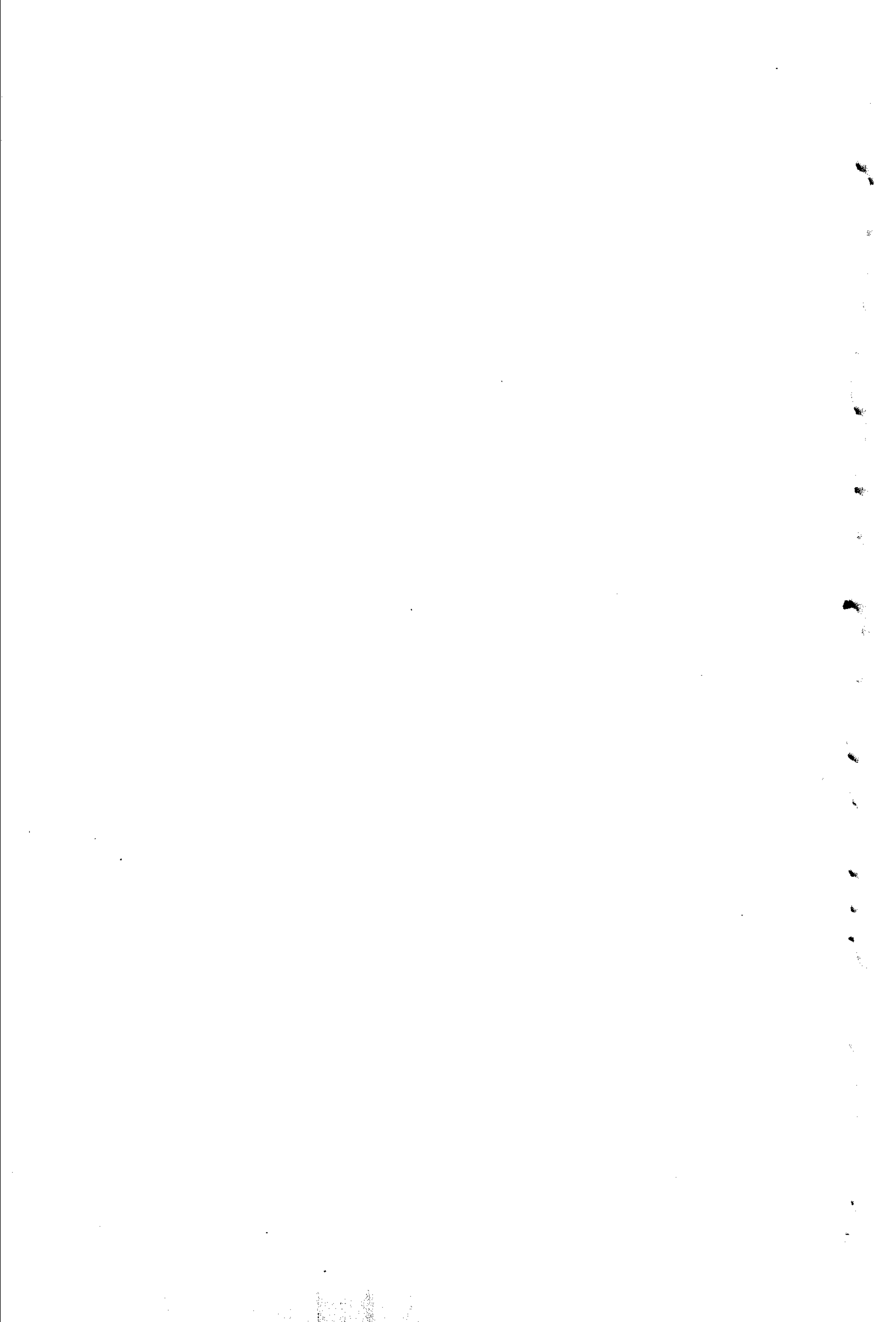
很容易看出，我们这里不用任何度量. 自然我们得到的应该是一个拓扑不变量. 这个 Feynman 积分依赖于两个变量，一个是作用量中的 k，另一个是群 $SU(N)$ 中的 N，这两个变量只能取整数值，Witten 证明了，解析开拓 k 和 N 到复变量给出两个变量的 Jones 多项式.

Witten 不仅给出 Jones 多项式的这个解释，早前，他有过一个 Morse 不等式的看法. 1988 年他还给出一个 4 维超对称规范场模型来解释 Donaldson 的多项式和 Floer 同调. 数学与物理在这里这样好地携起手来，实在是很让人振奋的.

低维拓扑学确实是近十几年来数学中最活跃的分支. Thurson 因为他在 2 维、3 维流形上杰出的工作得到了 1982 年的 Fields 奖，丘成桐的极小曲面的工作也是他得到 1982 年 Fields 奖的主要工作之一. Freedman 和 Donaldson 因为他们在 4 维流形上的工作得到了 1986 年的 Fields 奖，而 Jones 和 Witten 则因为他们各自的工作得到了 1990 年 Fields 奖. 特别是理论物理学作为一种工具或作为一种看法，与低维拓扑学还有什么关系，还有很多问题值得去研究. 很多大数学家仍在不断地思考. 作为现代数学主流中的一部分的低维拓扑学还会继续发展一个时期.

VI

复　分　析
Complex Analysis



杨　乐
（中国科学院数学研究所）
Yang Lo
（Institute of Mathematics,
Academia Sinica）
陈志华
（上海交通大学应用数学系）
Chen Zhihua
（Department of Applied Mathematics,
Shanghai Jiaotong University）

复　分　析
Complex Analysis

复分析是数学基础理论里的一个重要学科,同时它在其他数学分支以及空气动力学、弹性力学、理论物理等许多领域中有着广泛的应用.

作为"现代数学中若干基本问题的研究"这个科学基金重大项目的一个组成部分,项目"复分析"侧重于多复变函数论的研究.由于要求本文面向不专门从事数学研究与教学的广大读者,我们将由单复变函数开始进行论述,以便读者较好的了解.

§1 单复变函数

§1.1 引言

作为高等数学最基本内容的微积分,主要是讨论实变函数 $f(x)$ 的求导数与积分两大问题.导数与积分的定义不难推广到复变函数 $f(z)$ 的情形.

虽然复变函数与实变函数的导数定义十分相似,然而它们蕴含的实质却有很大不同.复变函数 $f(z)$ 是定义在一平面区域 G 内,它在 G 内点 z_0 存在导数是指变量 z 在 G 内沿着任意方向和路径趋向 z_0 时,$\dfrac{f(z)-f(z_0)}{z-z_0}$ 的极限总存在且具有相同的值.而在实变函数时,变量只能沿着实轴仅由两个方向趋向于定点.这就决定了在域 G 内每点可导的复函数(称为全纯函数或解析函数)$f(z)$ 具有一系列优秀的性质,因此形成了复分析这一独立的学科.

对于全纯函数,现在我们举述几条优秀性质[1].设 $f(z)$ 在域 G 内定义且在 G 内每点存在导数,则

(1) 它在 G 内必满足 Cauchy-Riemann 方程

$$\frac{\partial f}{\partial \bar{z}} = 0, \text{其中} \frac{\partial}{\partial \bar{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right);$$

(2)$f(z)$ 可以被 G 内它在一个点列上的值唯一确定，只要这个点列在 G 内有一个聚点；

(3)$f(z)$ 在 G 内必定无限次可导，且在 G 内任意点 z_0 的小邻域内 $f(z)$ 可以展开为$(z - z_0)$ 的收敛幂级数；

(4) 当 G 为开平面，即假定 $f(z)$ 在 $|z| < \infty$ 每点可导，如果设 $f(z)$ 在开平面有界，则它必蜕化为常数.

显然这些性质都是可微的实变函数所不具备的，以(4) 为例，$\sin x$ 与 $\cos x$ 都是在$(-\infty, +\infty)$ 上可微的函数，而且在此区间上它们的绝对值都不超过 1，但是它们并不恒等于常数.

复分析是一门历史悠久的学科. 如果从 19 世纪上半叶的 Cauchy 算起，迄今已超过一个半世纪，在漫长的岁月里，按照复变函数的不同研究主题，形成了一些不同的分支学科与研究领域. 现在我们介绍其中的几个问题，除去例外值问题，其余问题仅作十分简略的介绍.

§1.2 例外值问题

类似于求代数方程的根，一些理论与实际问题要求我们研究函数方程的根. 若 $f(z)$ 是定义在开平面 $|z| < \infty$ 上的全纯函数(通常称为整函数)，a 是一任意复数，问题是方程 $f(z) = a$ 有没有根?这些根的"数量"与分布如何?

复分析中的 Weierstrass 定理可以看作是对这个问题的最初的近似解答. 1879 年，法国著名数学家 E. Picard 证明了下述深入的结果：设 $f(z)$ 是整函数，不蜕化为常数，则对于任意的有穷复数 a，方程 $f(z) = a$ 恒有根，至多可能有一个复数例外，这个例外的复数就称为 $f(z)$ 的 Picard 例外值.

例如 e^z 是一整函数，对于任意的有穷非零复数 a，$e^z = a$ 有无

穷多个根 $\log|a| + i(\arg a + 2k\pi)(k = 0, \pm 1, \pm 2, \cdots)$,0 则是 e^z 的 Picard 例外值.

以后,É. Borel 引进了整函数的级的概念,对 Picard 定理作了定量的推广. 与这些定理相联系,E. Landau, F. Schottky, J. Hadamard, P. Montel, A. Denjoy, G. Julia, G. Valiron 等相继获得了一系列优秀的成果. Picard-Borel 定理不难推广到亚纯函数的情况. 开平面上的亚纯函数就是两个整函数的商,正如有理函数可以表为两个多项式的商. 1925 年,R. Nevanlinna 对亚纯函数的值分布建立了普遍的理论,使得这方面的研究进入了一个新时期.

设函数 $f(z)$ 于开平面亚纯,定义[41]

$$m(r,f) = \frac{1}{2\pi}\int_0^{2\pi} \log^+ |f(re^{i\theta})| d\theta,$$

$$N\left(r,\frac{1}{f-a}\right) = \int_0^r \frac{n\left(t,\frac{1}{f-a}\right)}{t} dt,$$

其中 $\log^+ \alpha = \max(\log\alpha, 0)$, $n\left(t,\frac{1}{f-a}\right)$ 为 $|z| \leqslant t$ 上 $f(z) = a$ 根的个数,重根须计及其重数. 当 a 为无穷时,$N\left(r,\frac{1}{f-a}\right)$ 记为 $N(r,f)$. Nevanlinna 引入了 $f(z)$ 的特征函数

$$T(r,f) = m(r,f) + N(r,f)$$

与

$$\delta(a,f) = 1 - \varlimsup_{r\to\infty} \frac{N\left(r,\frac{1}{f-a}\right)}{T(r,f)}.$$

当 $\delta(a,f) > 0$ 时,则 a 称为 $f(z)$ 的亏值,$\delta(a,f)$ 则称为亏量. Nevanlinna 理论的最主要结果是下述的亏量关系.

设 $f(z)$ 于开平面亚纯,则它至多只能有可数个亏值,且其相应的亏量总和不超过 2,即 $\sum_{a\in\bar{C}}\delta(a,f)\leqslant 2$.

Nevanlinna 的基本定理与亏量关系建立后,关于亏值与亏量

有很多研究工作，这里我们只稍微举述几个.

设 $f(z)$ 为开平面上的亚纯函数，α 为小于1的正数. 由于 $\delta(a, f) \leqslant 1$，因此 $\delta(a,f) \leqslant \delta(a,f)^{\alpha}$. 由亏量关系有 $\sum_{a\in\bar{C}}\delta(a,f)\leqslant 2$，但是 $\sum_{a\in\bar{C}}\delta(a,f)^{\alpha}$ 是否收敛呢？W. Fuchs[26] 首先证明了当 $f(z)$ 的级有穷*，$\alpha=\frac{1}{2}$ 时上述级数收敛. W. K. Hayman[30] 进一步证明了：若 $f(z)$ 于开平面亚纯，且级为有穷，则对任意 $\alpha>\frac{1}{3}$ 有 $\sum_{a\in\bar{C}}\delta(a, f)^{\alpha}<+\infty$；当 $\alpha<\frac{1}{3}$ 时，确实存在于开平面亚纯的有穷级函数使上述级数发散. 遗留下 $\alpha=\frac{1}{3}$ 的情况最终为 A. Weitsman[49] 证实，他只假定了亚纯函数 $f(z)$ 的下级有穷，证明了 $\sum_{a\in\bar{C}}\delta(a,f)^{\frac{1}{3}}$收敛.

关于亏值的另一个有趣的问题是所谓 Nevanlinna 反问题，即任意给定一列复数 $\{a_j\}$ 与一列正数 $\{\delta_j\}$ 适合 $\delta_j\leqslant 1$，$\sum\delta_j\leqslant 2$，是否存在开平面上亚纯函数 $f(z)$ 使得 $\delta(a_j,f)=\delta_j(j=1,2,\cdots)$ 以及对于其他任何复数 $a\neq a_j$ 有 $\delta(a,f)=0$？经过许多学者研究后，这个问题最终为 D. Drasin[19] 所解决. 除去值分布论，他主要使用了拟共形映照作为有效的工具. 由于有穷级亚纯函数 $f(z)$ 的亏量必须适合 $\sum_{a\in\bar{C}}\delta(a,f)^{\frac{1}{3}}<\infty$，所以在一般情况下 Nevanlinna 反问题中的亚纯函数必须是无穷级. 于是关于有穷级亚纯函数的反问题又被提了出来. 最近 A. E. Eremenko[24] 成功地证明了：任给一列复数 $\{a_j\}$ 与一列正数 $\{\delta_j\}$ 适合 $\delta_j<1$ 与 $\sum\delta_j\leqslant 2$，$\sum\delta_j^{\frac{1}{3}}$

---

* $f(z)$ 的级刻划了其特征函数的增长性，它的定义为 $\lambda=\varlimsup_{r\to+\infty}\frac{\log T(r,f)}{\log r}$；$f(z)$ 的下级则定义为 $\mu=\varliminf_{r\to+\infty}\frac{\log T(r,f)}{\log r}$.

$<+\infty$，则必存在于开平面亚纯的有穷级函数 $f(z)$ 使得 $\delta(a_j,f)=\delta_j(j=1,2,\cdots)$ 以及对于其他任意复数 a（即 $a\neq a_j$）有 $\delta(a,f)=0$.

1930 年前后，F. Nevanlinna 曾提出下述猜测：设函数 $f(z)$ 于开平面亚纯，下级 μ 为有穷，适合条件 $\sum_{a\in\overline{C}}\delta(a,f)=2$，则(1) $f(z)$ 的级 λ 应等于 μ，且为 $\frac{1}{2}$ 的正整数倍；(2) $f(z)$ 的每个亏量必为 $\frac{1}{\mu}$ 的整数倍；(3) $f(z)$ 的每个亏值必同时为其渐近值.（复数 a 称为是 $f(z)$ 的一个渐近值，是指存在伸展到无穷的连续曲线 Γ，当 z 沿着 Γ 趋向于无穷时，$f(z)$ 以 a 为极限.）当 $f(z)$ 是整函数时，情况较为简单，F. Nevanlinna 猜想早为 A. Pfluger[43]，A. Edrei 与 W. Fuchs[23] 证实. 当函数 $f(z)$ 亚纯时，问题要困难得多，最近已为 D. Drasin[20] 所彻底解决.

亚纯函数的亏值概念还可以发展为所谓亏函数. 设函数 $f(z)$ 于开平面亚纯，$a(z)$ 亦为开平面的任意一个亚纯函数，适合条件 $T(r,a(z))=o\{T(r,f)\}$. 类似于亏值与亏量的定义，可以引入 $f(z)$ 对于小函数 $a(z)$ 的亏函数与亏量的概念，这方面也有许多研究工作，特别应该提到的是 C. Osgood[42] 与 N. Steinmetz[47] 已经将亏量关系推广到亏函数的情况，即他们证明了

设 $f(z)$ 是开平面上的亚纯函数，则其亏函数至多是可数的，且相应的亏量总和不超过 2.

关于亏值的其他重要研究，诸如展布关系，亏值数目与辐角分布、亚纯函数的偏差系数等，我们就不在这里赘述了.* 这方面还有一些重要问题尚未解决. 例如对于级大于 1 的亚纯函数，其亏量总和是否可用更精确的上界去估计？在整函数的情况，则有所谓 Arakeljan 猜想[4,36] 还有待证实. 最近 Lewis 与 Wu[36] 的工作迈出

* 可参阅[16,31,52].

了可喜的第一步.*

§1.3 唯一性问题

在值分布论中,有一个较易为人们理解的问题 —— 唯一性问题.

对于函数 $f(z)$ 与复数 a,用 $E(f=a)$ 表示 $f(z)=a$ 的全体根构成的集合.容易看出一个多项多 $P(z)$ 可以被两个集合 $E(p=0)$ 与 $E(p=1)$(或者一般地 $E(p=a)$ 与 $E(p=b)$,其中 a 与 b 为两个判别的有穷复数)完全确定.有理函数 $R(z)$ 则可以被三个集合 $E(R=0)$,$E(R=\infty)$ 与 $E(R=1)$ 完全确定.于是,人们自然会问,一个亚纯函数 $f(z)$ 需要几个这样的集合 $E(f=a)$ 才能唯一确定呢?这就是亚纯函数的唯一性问题.

应用 Nevanlinna 的基本定理不难证明开平面上的亚纯函数 $f(z)$ 可以被 5 个集合 $E(f=a_j)(j=1,2,\cdots,5)$ 完全确定,这里 $a_j(j=1,2,\cdots,5)$ 是 5 个互相判别的复数.此结果中的 5 是最佳的,事实上亚纯函数 e^z 与 e^{-z} 有 4 个相同的取值集合.即 $E(e^z=0)\equiv E(e^{-z}=0)=\varnothing$,$E(e^z=\infty)\equiv E(e^{-z}=\infty)=\varnothing$,$E(e^z=1)\equiv E(e^{-z}=1)$ 与 $E(e^z=-1)\equiv E(e^{-z}=-1)$.因此,由 4 个取值集合不能唯一确定亚纯函数.

关于整函数与亚纯函数的唯一性问题,也有不少的讨论.例如作为 Osgood 与 Steinmetz 关于小函数的基本不等式的应用,立即有:于开平面上亚纯的函数 $f(z)$ 可以被 5 个集合 $E(f=a_j(z))(j=1,2,\cdots,5)$ 完全确定,这里 $a_j(z)(j=1,2,\cdots,5)$ 是 5 个互相判别的亚纯函数,且适合条件 $T(r,a_j(z))=o\{T(r,f)\}$.关于例外值问题和唯一性问题的许多结果已被推广到代数体函数、全纯

---

* 此文写成后,笔者得知 A. E. Eremenko 于 1991 年 12 月构造出例子说明 Arakeljan 猜想不成立.当然还可以考虑较 Arakeljan 猜想弱的结果是否成立.

与亚纯曲线以及多复变函数的情形.

§1.4 单叶函数的系数估计——Bieberbach 猜想

长期以来,几何函数论研究里的一个中心课题是要证实 Bieberbach 猜想.它已于 1984 年为 Louis de Brange[10] 彻底解决.

设 $f(z)$ 于 $|z|<1$ 内全纯,并且单叶,即若 z_1 与 z_2 是 $|z|<1$ 内任意两个不同的点,则有 $f(z_1)\neq f(z_2)$. 若再假定 $f(z)$ 适合就范性条件 $f(0)=0, f'(0)=1$,则有展式 $f(z)=z+\sum\limits_{n=2}^{\infty}a_nz^n$. 通常将满足上述条件的函数类记为 S. 1916 年,L. Bieberbach 证明了若 $f(z)\in S$,则 有 $|a_2|\leqslant 2$ 以后,提出了一般性的猜想 $|a_n|\leqslant n(n=2,3,4,\cdots)$. 由于函数 $K(z)=\dfrac{1}{(1-z)^2}=\sum\limits_{n=1}^{\infty}nz^n$ 属于 S,所以猜想中的系数上界是精确的.

从 1916 年至 80 年代初的漫长岁月里,围绕着 Bieberbach 猜想有大量的研究工作[22],形成了一些重要的方法,也获得了许多显著的结果. 诸如参数表示法、变分方法、面积原理的推广、Grunsky 不等式以及 FitzGerald 不等式等.

应用了参数表示法、Milin-Lebedev 不等式,Louis de Brange 巧妙地构造了一组辅助函数,将 Bieberbach 猜想化为特殊函数的一已知结果,从而完全证实了这个猜想.

§1.5 正规族与 Bloch 法则

以上的几个问题只是对每个函数作单独的考察,然而有的复分析问题是要研究一族函数所具有的性质. 正规族就是其中简单而重要的一种.

区域 G 内的一族全纯函数(或亚纯函数) $\mathscr{F}$ 称为是 G 内的一正规族,如果 $\mathscr{F}$ 的任意一函数序列必定包含一个子序列在 G 内任

一闭域上按球面距离一致收敛.

正规族是一个十分有力的工具,常被用于存在性定理的证明之中. 于是判定一个函数族是否正规就成了重要问题. 有趣的是这个问题又与函数取值的问题紧密地联系在一起. 试看下述 Montel 的正规定则.

设 $\mathscr{F}$ 是域 G 内的一族全纯函数,a 与 b 是两个判别的有穷复数. 若 $\mathscr{F}$ 中的每一函数 $f(z)$ 在 G 内均不取 a 与 b,则 $\mathscr{F}$ 在 G 内是正规的.

比较关于函数取值的 Picard 定理(见 §1.2) 与上述 Montel 的正规定则,A. Bloch 产生了一个奇妙的想法:如果存在某个(或若干个) 条件 P,具有下述性质,在开平面全纯(或亚纯) 的函数 $f(z)$ 适合条件 P,则 $f(z)$ 必蜕化为常数. 若在域 G 内有一全纯(或亚纯) 函数族 $\mathscr{F}$,$\mathscr{F}$ 中每一函数 $f(z)$ 在 G 内都一致地适合条件 P,则 $\mathscr{F}$ 在 G 内应为正规族. 简单地说,对于每一个 Picard 型的定理,就相应地有一个 Montel 型的正规定则. Bloch 的这个思想后来被称为 Bloch 法则.

虽然可以找出例子说明 Bloch 法则并不永远成立,然而涉及到值分布的问题它却常常是正确的. 因此,Bloch 法则是预言新正规定则的有力武器[31]. 这里我们举一个最近的例子. *

J. Clunie[17] 早就证明:若 $f(z)$ 是整函数,a 是一非零复数,且 $f(z)f'(z)\neq a$,则 $f(z)$ 必蜕化为常数. 按照 Bloch 法则,人们猜想下述正规定则应该成立. 设 F 是域 G 内的一族全纯函数,a 是一非零有穷复数. 若对于 $\mathscr{F}$ 中每个函数 $f(z)$ 有 $f(z)f'(z)\neq a$,则 $\mathscr{F}$ 在 G 内为正规族. 这个正规定则已于最近获得证实[44].

---

* 其他的例子可参阅[31,51,52].

§ 1.6 调和测度与 Hausdroff 测度

在复分析中，有一系列关于点集度量的概念. 例如解析容量、超越直径、调和测度以及 Hausdroff 测度等，这些都是复变函数论中十分有效的工具. 于此，我们介绍一下调和测度与 Hausdroff 测度.

设域 G 的边界 ∂G 上有一点集 α，它由有限个弧段组成. 在 G 上存在唯一的调和函数，在 α 上每点等于 1，而在 $\partial G-\alpha$ 上等于零. 它在 G 内点 z 的值，称为 α 对于域 G 在点 z 的调和测度，记为 $\omega(z,\alpha,G)$. 调和测度的概念可以推广到平面上的任意点集[41]. 应用调和测度这一有力工具，不难导出关于全纯函数最大模的二常数定理、Hadamard 的三圆定理以及关于整函数渐近值数目的 Denjoy 猜想.

设 α 是一平面点集，作为线性测度、平面测度的推广，有所谓 Hausdroff 测度[41]. 设 $h(r)$ 是在 $r \geqslant 0$ 时定义的连续递增的函数，$h(0)=0$. 用任意一列圆 C_j 来覆盖点集 α，每个圆 C_j 的半径 $r_j \leqslant \varepsilon$. 对于所有可能的这种覆盖，相应的 $\sum h(r_j)$ 的下确界记为 $m_\varepsilon(\alpha, h)$. 再令 ε 趋向于 0，$\lim\limits_{\varepsilon\to 0} m_\varepsilon(\alpha,h)=m(\alpha,h)$ 则称为点集 α 相对于 $h(r)$ 的 Hausdroff 测度.

那么，给定一个平面点集 E，它的调和测度与 Hausdroff 测度之间的关系是什么?这是一个重要而令人关注的问题，曾经有不少的研究工作. 1985 年苏联青年学者 N. G. Makarov[37] 成功地解决了这个问题.

§ 1.7 全纯函数与亚纯函数的边界性质

在 § 1.1 中我们已经看到，域 G 内的全纯函数 $f(z)$ 具有许多优秀的性质. 这是因为对于复变函数来说可导的性质是一个很强

的条件. 尽管如此，当点 z 趋向于域 G 的边界点时，情况就复杂得多了. 于是研究全纯函数与亚纯函数在边界附近的状态就成了一个重要的研究课题. 这方面的基本结果是 Fatou 定理[18].

设函数 $f(z)$ 在单位圆内全纯且有界，Fatou 定理说 $f(z)$ 的径向边界值必定几乎处处存在. 结论中的径向边界值还可以改为角形边界值，即趋向于边界时不沿着与边界相切的路径.

关于全纯函数与亚纯函数的边界性质，有一系列重要的研究工作，例如下述有趣的 Plesner 定理[18].

设 $f(z)$ 于单位圆内亚纯. 边界点 $e^{i\theta_0}$ 称为 $f(z)$ 的一个 Fatou 点，如果 $f(z)$ 在 $e^{i\theta_0}$ 存在角形边界值. $e^{i\theta_0}$ 称为 $f(z)$ 的 Weierstrass 点，如果 $f(z)$ 把以 $e^{i\theta_0}$ 为顶点的任一圆周角映为复平面上处处稠密的集合. 在此定义下，Plesner 定理说：若 $f(z)$ 在单位圆内亚纯，则所有的边界点可能除去一零测度集外由 Fatou 点与 Weierstrass 点构成.

§ 1.8 插值问题

在单位圆内全纯的函数 $f(z)$ 称为属于 $H^p(0<p<+\infty)$ 空间，如果

$$\sup_{r<1}\left\{\int_0^{2\pi}|f(re^{i\theta})|^p d\theta\right\}<+\infty.$$

设 $\{z_k\}$ 是 $|z|<1$ 内的任意一点列，$\{w_k\}$ 为任意一列复数，插值问题是讨论在什么样的情况下存在函数 $f(z)\in H^p$ 使得 $f(z_k)=w_k, k=1,2,\cdots$.

为了叙述插值的一个重要结果，我们先引入一致分离的概念[21]. $|z|<1$ 内的点列 $\{z_k\}$ 称为一致分离，如果存在正数 δ 使

$$\prod_{\substack{j=1\\ j\neq k}}^{\infty}\left|\frac{z_k-z_j}{1-\bar{z}_j z_k}\right|\geqslant\delta,\quad k=1,2,\cdots.$$

再记 T_p 为 $H^p(0<p\leqslant\infty$，H^∞ 为 $|z|<1$ 上全体有界全纯函数

构成的空间）上的线性算子，由

$$T_p(f) = \left\{(1 - |z_k|^2)^{\frac{1}{p}} f(z_k)\right\}$$

定义. 记 $l^p(0 < p < \infty)$ 为以数列 $(a_1, a_2, \cdots)$ 为元素构成的空间，适合条件 $\sum |a_j|^p < +\infty$. l^∞ 则为全体有界数列构成的空间. 在这些定义下，L. Carleson[12] 证明了：若 $0 < p \leqslant \infty$，则 $T_p(H^p) = l^p$ 的充要条件是 $\{z_k\}$ 为一致分离的. 在证明中，他引进了以后被称为Carleson 测度的重要概念，Carleson 测度在 H^p 空间的一些其他方面也起了很大作用.

以上，我们只是在单复变函数论中挑选了几个容易为广大读者了解的问题作了简略介绍. 实际上，还有很多非常重要的方面，诸如 Riemann-Roch 定理，拟共形映照，Teichmüller 空间，复动力系统，等等，这里就不再涉及了.

§2　多复变函数

§2.1　概论

多复变函数论是复分析中的一个重要分支，近年来有较快的发展，据粗略统计在过去 20 年间它的研究论文增加了 9 倍. 多复变函数论近年来受到不少数学家的重视是因为多复变函数论研究最能体现数学的一致性. 多复变函数论与代数，代数几何，微分几何，偏微分方程，泛函分析，代数拓扑，实分析之间的丰硕的相互作用，不仅促进了多复变函数论的研究，亦导致、刺激这些学科的重要发展.

多复变函数论的最早研究，自然是研究多个复变数的空间 $\mathbf{C}^n$ 中的区域上的全纯函数. 这里 $z = (z_1, \cdots, z_n) \in \mathbf{C}^n$, $z_1, \cdots, z_n$ 都

表示是一个复数. 设 $\mathscr{D}$ 是 C^n 中的一个区域, $\mathscr{D}$ 上的一个复值函数 $f: \mathscr{D}\to C$, $z\mapsto w=f(z)$. 第一件事就是什么样的 f 是全纯函数？令 $B(a,r)=\left\{z\in C^n \mid |z-a|=\left(\sum |z_i-a_i|^2\right)^{\frac{1}{2}}<r\right\}$ 表示以 $a=(a_1,\cdots,a_n)\in C^n$ 为中心, $r>0$ 为半径的球, 对 $a\in\mathscr{D}$, 如存在 $B(a,r)\subset\mathscr{D}$, 使 f 在 $B(a,r)$ 内能展开为一致收敛的幂级数

$$f(z)=\sum_{\substack{0\leqslant\alpha_i\\1\leqslant i\leqslant n}} c_{\alpha_1\cdots\alpha_n}(z_1-a_1)^{\alpha_1}\cdots(z_n-a_n)^{\alpha_n},$$

就称为 f 在 a 点是全纯的. 如果对 $\mathscr{D}$ 的每一点 f 都是全纯的, 就称 f 是 $\mathscr{D}$ 上的全纯函数. 这个定义是单复变全纯函数的推广. 众所周知单复变全纯函数还有一个常用的定义就是 Cauchy-Riemann 方程, 在单复变数中这两个定义的等价是一件易于了解的事实. 作为多复变函数来讲这就是古典的 Hartogs 定理, $f: \mathscr{D}\to C$ 的复值函数, 如果 f 对每个变数 z_i, $1\leqslant i\leqslant n$ 是全纯的, 即 f 对每个 z_i 适合 Cauchy-Riemann 方程

$$\frac{\partial f}{\partial\bar{z}_i}=0,\ \frac{\partial}{\partial\bar{z}_i}=\frac{1}{2}\left(\frac{\partial}{\partial x_i}+\sqrt{-1}\,\frac{\partial}{\partial y_i}\right),\quad z_i=x_i+\sqrt{-1}\,y_i,$$

则 f 就是全纯的. Hartogs 定理的证明是很富有技巧性的, 读者有兴趣可参考[40]. Hartogs 定理使我们对多复变函数的全纯函数的了解大为方便, 现在大家都将对每个自变数都是全纯的函数作为多复变函数的全纯函数.

无疑从多复变数的全纯函数的定义出发, 我们可以得到一些与单复变数全纯函数类似的性质. 但是多复变函数论作为一个重要的数学分支却恰恰是它所呈现的与单复变函数论完全不同的方面, 下面就从简单易懂的角度来逐个介绍.

§2.2 全纯域与拟凸域

对单复变数来讲, C 平面上的任一区域 $\mathscr{D}$ 上一定存在有全纯函数无法开拓到 $\mathscr{D}$ 外更大区域上. 但是在 C^n 中这一事实一般不

再成立. 1906年Hartogs发现C^n中存在这样的区域G,任何G上的全纯函数均可以开拓到更大的区域上去,这种区域的存在后来被人们称为是Hartogs现象. 具有Hartogs现象的区域很多,例如$\mathscr{D}$是C^n中的任一个区域,对$\forall\ a \in \mathscr{D}$与$\overline{B(a,r)} \subset \mathscr{D}$,则$G = \mathscr{D} - \overline{B(a,r)}$就具有Hartogs现象,即任何一个$G$上全纯函数,一定可以开拓到$\mathscr{D}$上去. 另一个简单例子是$G = \{(z,w) \in C^2 \mid |z| < 1, \beta < |w| < 1; \beta < 1\} \cup \{(z,w \in C^2 \mid |z| < \alpha < 1, |w| < 1\}$,则每个$G$上的全纯函数,一定可以开拓到双圆柱$\{(z,w) \mid |z| < 1, |w| < 1\}$.

对C^n中不具有Hartogs现象的区域,就称之为全纯域,无疑全纯域是更为有价值的区域,因为它的边界是自然边界,即若G是全纯域,则存在G上的全纯函数f,它的自然定义域就是G. 关于刻划全纯域还有一个方法就是全纯凸性. 今用$\mathscr{I}(G)$表示G上的全纯函数的全体,设K是G中的紧集,

$$\hat{K} = \{z \in G \mid |f(z)| \leqslant |f|_K \quad \forall\ f \in \mathscr{I}(G)\},$$

$\hat{K}$称为K(关于G)的全纯凸包,如对G中任一紧集K,$\hat{K}$都是G中的紧集,就称G是全纯凸域. 易于证明全纯凸域就是全纯域.

在上面提到的全纯域时并没有涉及这个区域的边界. 在多复变函数中,区域的边界具有较复杂的结构,并不像单复变那样区域的边界均是由曲线组成. 例如前面提到的双圆柱$P_2 = \{(z,w) \in C^2 \mid |z| < 1, |w| < 1\}$,它的边界是由三部分组成,它分别是$\{(z,w) \in C^2 \mid |z| = 1, |w| < 1\}$,$\{(z,w) \in C^2 \mid |z| < 1, |w| = 1\}$与$\{(z,w) \in C^2 \mid |z| = 1, |w| = 1\}$,这中间前两部分都是实3维的,最后一部分是实2维的. 应该讲最有兴趣,亦是边界情况较为简单的C^n中的区域,是它的边界是由$2n-1$维光滑曲面围成.

设G是C^n中有界区域,而且其边界∂G是由可微曲面($2n-1$维)组成. 今设$\partial G = \{z \in C^n \mid r(z) = 0\}$和$G = \{z \in C^n \mid r(z) <$

$0\}$，这里 $r(z)$ 是在 ∂G 的邻域中可微的，而且 $\mathrm{grad}\ r(z)|\partial G$ 无处为 0. 1910 年 E. E. Levi 提出对于这样的区域，能否用其边界 ∂G 的性质来刻划 G 的全纯凸性.

定义　设 G 如上给定，如果 G 上定义函数 r 的 Levi 形式 $Lr = \sum_{\alpha,\beta} \frac{\partial^2 r}{\partial z_\alpha \partial \bar{z}_\beta} dz^\alpha \otimes \overline{dz^\beta}$ 限制在 ∂G 的复切平面上，即 $\forall\ x \in \partial G$，使得

$$\sum_{\alpha,\beta} \frac{\partial^2 r}{\partial z_\alpha \partial \bar{z}_\beta}(x)\xi^\alpha \overline{\xi^\beta} \geqslant 0 \qquad (\star)$$

对 $\forall\ (\xi^1, \cdots, \xi^n) \in C^n \quad (\xi^1, \cdots, \xi^n) \neq 0$ 与 $\sum_{\alpha=1}^{n} \frac{\partial r}{\partial z^\alpha}\xi^\alpha = 0$，就称 G 是拟凸域，如果($\star$)式是正的，则称 G 为强拟凸域.

可以证明每个拟凸域都是全纯凸的，因此拟凸域也是全纯域. 大家易于发现最简单的强拟凸域就是 C^n 中的单位球 $B(0,1)$，此时只要取 $r = \sum_{i=1}^{n} |z_i|^2 - 1$ 即可. 如果 G 是以 C^2 边界 ∂G 的拟凸域，则对 $\forall\ z \in G$，定义 $d(z) = \mathrm{dist}(z, \partial G)$，可以验证实值函数 $r = -\log d(z)$，其 Levi 形式 Lr 是在 G 上正定的，这样的 r 称为 G 上的强多次调和函数. 另外对 $\forall\ c \in R$，$G_c = \{z \in G | r(z) \leqslant c\}$ 是 G 中的紧集，具有这样性质的函数称为 G 上的穷竭函数. 所以在拟凸域上存在有一个强多次调和穷竭函数.

拟凸域有很多重要的性质，例如单复变数的 Mittag-Leffler 定理与除法定理都在拟凸域上成立. 关于拟凸域已有很多研究结果，但还有很多重要问题值得研究，这在以下将会有所涉及.

§2.3　全纯映射

作为多复变函数单是讨论 C^n 中的区域 G 上的全纯函数自然是不够的，显然应该讨论 $F = (f_1, \cdots, f_m): G \to C^m$ 的全纯映射，亦即这里 F 中的每一个元素 $f_i (1 \leqslant i \leqslant m)$ 都是 G 上的全纯函数，

其中尤以讨论 $m = n$ 的情况具有特殊的意义.

关于全纯映射第一个重要问题就是 C^n 的区域之间的全纯等价问题,亦即如 G_1,G_2 都是 C^n 的区域,何时存在全纯映射 $F:G_1 \to G_2$ 与 $H:G_2 \to G_1$,而且 $H \circ F:G_1 \to G_1$ 与 $F \circ H:G_2 \to G_2$ 都是恒同映射. 在单复变中有一个著名的 Riemann 映照定理;复平面上边界多于一点的单连通区域一定全纯等价于单位圆盘. 而在多复变函数中域之间的全纯等价问题确是极其复杂的问题,就连 C^2中的单位球 $B(0,1)$ 与双圆柱 $P_2 = \{(z_1,z_2) \in C^2 \mid |z_1| < 1, |z_2| < 1\}$ 这两个在 C^2中最简单的单连通区域,彼此都不全纯等价.

关于多复变函数的区域全纯等价方面最为完整的结果是关于有界对称域的分类(在全纯等价下). 一个区域 G 到其自身的可逆全纯映射,称为 G 上的一个全纯变换,所有 G 上全纯变换关于变换的复合构成一个群,称为 G 的全纯变换群,一般用 $\mathrm{Aut}(G)$ 表示. 如果对每个 $p \in G$,均存在 $F_p \in \mathrm{Aut}(G)$,使 p 为 F_p 的唯一孤立点,而且 $F_p \circ F_p =$ 恒等映射,就称这个 G 是对称的. É. Cartan 用 Lie 代数的方法对有界对称域进行了完全的分类,在此基础上华罗庚教授对这些分类的代表域给出了明确的矩阵表示,这些域称之为典型域. 因为这些区域有明确的矩阵表示,华罗庚教授在其上讨论了调和分析[35]. 从有界对称域的定义知道,这类区域具有极强的对称性,它的分类完全依赖这样的对称性. 比有界对称域更广一些的区域就是可递域,即是对 $\forall\ p_1,p_2 \in G$,一定存在 $F \in \mathrm{Aut}(G)$,使 $F(p_1) = p_2$ 的区域. 继 E. Cartan 的工作之后,有界可递域的分类在过去 30 余年吸引了不少数学家去研究,但是至今进展不大,离开完全解决还很远. D. Burns, S. Shnider 和 R. Wells[11] 证明了单位球经过微小振动以后可以产生非可数个互不全纯等价的区域,特别是其中还有僵域(即指 $\mathrm{Aut}(G) = id_G$). 这个结果表明有界域的分类是十分复杂的.

对于 C^n 中全纯映射还有一个重要的问题就是边界对应问题，即如 $F:G_1 \to G_2$ 是一个双全纯映射，什么条件下，F 可以扩充为 $\overline{G_1}$ 到 $\overline{G_2}$ 的 C^∞ 微分同胚. 关于这个问题已知的最精彩的结果是 C. Fefferman[25] 的结果：若 G_1, G_2 都是 C^n 中的有光滑边界的强拟凸域，$F:G_1 \to G_2$ 是双全纯映射，则 F 可以扩充为 $\overline{\Omega}_1$ 到 $\overline{\Omega}_2$ 上的 C^∞ 微分同胚. Fefferman 这个定理，后来由 Bell 和 Ligoska[7] 所简化. Fefferman 定理导致了对更一般区域的讨论，现在已证明当 G_1 与 G_2 是有界拟凸区域并满足一个被称为条件 R 且 G_1 与 G_2 具有光滑边界时上述扩充定理成立，另外如果 G_1 与 G_2 是具有实解析边界的有界域时上述定理亦成立. 现在一般猜想对具光滑边界的有界域这个扩充定理也应该成立，但是对此并没有进一步的结果. S. S. Chern 和 J. Moser[15] 的论文讨论了 C^n 中的超曲面在全纯变换下的全系不变量，这些结果将会有助于对上面的问题的讨论与解决.

对于全纯映射另一个研究方面就是值分布理论，单复变亚纯函数的值分布理论是 20 世纪数学上最精细的，最巨大的成果之一. 在过去 30 年多复变函数的值分布理论亦有很大的发展. 借助于 L. V. Ahlfors 的关于一个变数由几何语言描绘的特征函数 $T_0(f,r)$ 的思想来建立相应的特征函数. 关于多复变函数值分布理论最早的论文是 S. S. Chern[14] 给出的，在那篇论文中建立了多复变函数全纯映射的第一基本定理. 在这之后值分布定理有很多发展，这里我们仅举述一两个简单的结果. 对于从 C^n 到 CP^n 的全纯映射，J. Carlson 和 P. Griffiths[13] 证明了亏量公式 $\sum_i \delta(Ai) \leqslant n + 1 = x(p^n)$，其中 A_i 是在 CP^n 中居最广位置的超平面. 关于 f: $C^1 \to CP^n$ 的全纯映射，当 $f(C)$ 全纯曲线与 $2n+1$ 居最广位置的 CP^n 的超平面无交时，则 f 一定是常数；当 $f(C)$ 与 $n+1$ 个居最广位置的超平面无交时，f 一定是退化的（即在 CP^n 的超平面之

中). 关于多复变函数的值分布的研究结果，读者可以参阅 W. Stoll[48] 的综合文章，对比单复变函数的丰富与精细的理论，读者会发现多复变的值分布理论尚有大量问题有待研究.

§2.4 复流形，Stein 流形

流形概念的引入是 20 世纪数学研究中的极为重要的发展，一个微分流形粗略的讲是一个 T_2 拓扑空间，它局部是 R^n 中的区域，而复流形是指它局部是 C^n 中的区域. 它的严格定义如下：

定义 设 M 是一个连通的 T_2 拓扑空间，设 $\{U_\alpha\}_{\alpha\in I}$ 是 M 的一个开覆盖，而且对每个 $\alpha\in I$，有 $\varphi_\alpha: U_\alpha\to\varphi_\alpha(U_\alpha)\subset C^n$ 是从 U_α 到 C^n 的开集的拓扑同胚. 当 $U_\alpha\cap U_\beta\neq\varnothing$ 时，$\varphi_\alpha\circ\varphi_\beta^{-1}:\varphi_\beta(U_\alpha\cap U_\beta)\to\varphi_\alpha(U_\alpha\cap U_\beta)$ 是 C^n 开集之间的全纯映射，则称 M 是一个复流形，n 称为 M 的复维数.

对点 $p\in U_\alpha$，$\varphi_\alpha(p)=(z_1^{(\alpha)},\cdots z_n^{(\alpha)})$ 就称为 p 点的局部坐标.

当 $n=1$ 时，M 就常称之为 Riemann 曲面，当 M 紧时，就称 M 为紧复流形.

关于复流形上的全纯函数或是复流形之间的全纯映射都是指它们局部坐标之间是全纯依赖关系.

关于复流形的研究首先值得提及的是 20 世纪 50 年代前后关于 Stein 流形的研究. 所谓 Stein 流形实际上就是前面提到的全纯域的推广.

定义 一个复流形 M 称为是一个 Stein 流形，如果它适合下面的条件：

(1)M 是全纯凸的；

(2)M 是全纯分离的，即对 $\forall\ p,q\in M,p\neq q$ 必存在 $f\in\mathcal{O}(M)$，使 $f(p)\neq f(q)$；

(3) 对 $\forall\ p \in M$ 均存在 $f_1, \cdots f_n \in \mathscr{O}(M)$，使 $(f_1, \cdots, f_m)$ 成为 p 点的局部坐标.

这里 $\mathscr{O}(M)$ 是 M 上的全纯函数空间.

由上述定义表示出 Stein 流形是具有很多全纯函数的复流形，从函数论观点来看这种流形有极大的重要性，另外全纯凸性也反映了流形具有全纯凸域那样的好性质. 关于 Stein 流形研究有两个最重要的结果，这就是著名的 Cartan 定理 A 与 Cartan 定理 B. Cartan 定理 A：如 M 是一个 Stein 流形，则对 M 上的任一解析凝聚层 φ 与 $\forall\ x \in M$，φ_x 由 $H^0(x, \varphi)$ 生成. Cartan 定理 B：如 M 是一个 Stein 流形，φ 是 M 上的解析凝聚层，则 $H^q(M, \varphi) = 0, q = 1, 2, 3, \cdots$. 这里 $H^q(M, \varphi)$ 是指以 φ 为系数的 q 阶上同调群. 当然 Cartan 定理 A，B 对更广的 Stein 空间亦成立. 因为篇幅的关系，本文不准备介绍复空间等概念. 在 Cartan 定理 A 与 B 中，Cartan 定理 B 更为本质，因为它可以蕴涵 Cartan 定理 A，而且一个复流形如果 Cartan 定理 B 成立，则它必是一个 Stein 流形. 如同全纯域一样 Stein 流形上的 Cousin 问题 I 有解（Cousin 问题 I 是单复变 Mittag-Leffler 定理的推广）. Cousin 问题 II 有解的一个充分条件是 $H^2(M, Z) = 0$（Cousin 问题 II 是单复变中 Weierstrass 定理的推广）. 由于 Stein 流形有较重要的性质，因此判定一个复流形何时是 Stein 流形成为大家感兴趣的课题. H. Grauert 的定理[27] 对此提供了有力的工具，H. Grauert 证明，如果一个复流形具有一个光滑的强多次调和穷竭函数，则这个流形一定是 Stein 流形. 因为在一个复流形上寻找这样的函数往往比验证 Stein 流形定义中三个条件远为简单与可行. R. E. Greene 与 H. Wu[28,29] 利用 Grauert 定理证明了截曲率为正的完备 Kähler 流形与截曲率为非正的单连通完备 Kähler 流形是 Stein 流形. 关于用曲率来判定 Stein 流形尚有很多问题有待探讨，例如具正双截曲率的非紧完备的 Kähler 流形是否是 Stein 流形？同样，一个负的双截曲率的完备单连通的

Kähler 流形是否是一个 Stein 流形？这些问题既未找到证明，亦未见有反例. 作为上面 Grauert 定理的推广，H. Grauert 与 A. Andreotti[3] 研究了复流形上的 q 强凸函数. 设 M 是一个 n 维复流形，$f: M \to R$ 是 M 上光滑的实值函数，如果 f 的 Levi 形式在一个紧集 K 外至少有 $n-q+1$ 个正特征根就称 f 为在紧集 K 外 q 强凸的. 一个复流形如果具有一个紧集 K 外 q 强凸穷竭函数，就称这个复流形是 q 凸的，如果 $K=\varnothing$，就称这个复流形是 q 完备的. A. Andreotti 和 H. Grauert 在 q 凸流形与 q 完备流形上推广了 Cartan 定理 B，他们证明，如复流形 M 是 q 凸的，则对任何解析凝聚层 φ，有 $\dim_C H^k(M,\varphi) \leqslant +\infty, k \geqslant q$；如复流形是 q 完备的，则有 $H^k(M,\varphi)=0, k \geqslant q$. 显然 q 完备就是 Stein 流形，近年来 H. Wu 有不少论文讨论 Kähler 流形的几何条件与 q 完备流形之间的关系，在此就不一一枚举了. 但是当 M 是一个 q 完备流形或 q 凸流形时，它的全纯函数空间的性质还所知甚少.

关于复流形的另一个研究方面就是单值化定理. 对于一维复流形，即 Riemann 曲面，著名单值化定理断言每个单连通的 Riemann 曲面必双全纯等价于单位圆盘，CP^1（即球面 S^2）与 C. 这其中 CP^1 是紧的，它上面有一个 Fubini-Study 度量，使它的 Gauss 曲率恒等于 $+1$，而 C 上普通的欧氏度量，它的 Gauss 曲率总为 0，单位圆盘上有一个 Poincaré 度量，它的 Gauss 曲率恒为 -1. 多复变函数的单值化定理是寻找复流形的曲率条件使其双全纯等价于有界域，CP^n 与 C^n. 这里关于双曲的情况是有界域而不是单位球，这是因为前面我们介绍过的有界域双全纯等价类本身就是十分复杂，因此无法在这里要求单值化定理的对象是单位球. 在这方面 Mori[38] 和 Siu-Yau[45] 证明了 Frankel 猜想，即双截曲率严格正的紧 Kähler 流形必双全纯同胚于 CP^n，对于抛物的情况 Siu-Yau[46] 证明了单连通完备 Kähler 流形，如它的截曲率满足 $0 \geqslant$ 截曲率

$\geqslant -\frac{A}{r^{2+\epsilon}}$，则其必双全纯等价于 C^n，这里 r 是到一定点的距离，A，ϵ 是两个正常数，而对于剩下的双曲情况至今没有满意的答案. 现在大家第一步有兴趣的是在什么样的几何条件下，一个非紧复流形可以具有有界全纯函数. 因为具有有界全纯函数是全纯等价于有界域的复流形的必要条件，但是至今这方面没有令人满意的结果.

§2.5 CR 子流形与 CR 扩充

一个复数，可以视作复平面上一个向量，复数乘上虚数单位 i，其几何意义就是旋转 $\frac{\pi}{2}$（正向），现在将复平面看成一个 2 维的实向量空间，将乘上虚数单位 i，命之为一个变换 J，则 $J: R^2 \to R^2$，而且 $J^2 = -id$. 现在将这个变换推广到 C^n 中去，设坐标 $z_1, \cdots, z_n$ 之对应实坐标为 $z_\alpha = x_\alpha + \sqrt{-1}y_\alpha$，用 $\frac{\partial}{\partial x_\alpha} \frac{\partial}{\partial y_\alpha}$，$\alpha = 1, \cdots, n$，分别表示 x_α, y_α 方向的单位向量，则定义 R^{2n} 上一个实线性变换 J，$J\left(\frac{\partial}{\partial x_\alpha}\right) = \frac{\partial}{\partial y_\alpha}$，$J\left(\frac{\partial}{\partial y_\alpha}\right) = -\frac{\partial}{\partial x_\alpha}$，$\alpha = 1, \cdots, n$，即对 R^{2n} 中任一个向量 $V = \sum_{\alpha=1}^{n}\left(a_\alpha \frac{\partial}{\partial x_\alpha} + b_\alpha \frac{\partial}{\partial y_\alpha}\right)$，$JV = \sum_{\alpha=1}^{n}\left(a_\alpha \frac{\partial}{\partial y_\alpha} - b_\alpha \frac{\partial}{\partial x_\alpha}\right)$，显然 $J^2 = -id$.

现在设 M 是 $C^n(R^{2n})$ 中的一个实 s 维实微分子流形，今对 $p \in M$，习知 $T_p(M)$ 表示 p 点切空间，实际上就是一个 s 维实平面，$T_p(M)$ 中的每个元素称之为 p 点的切向量，它也可视作是 $C^n(R^{2n})$ 中的向量. $JT_p(M) \cap T_p(M)$ 是一个偶维的 $T_\beta(M)$ 的子空间，今记它的维数为 $2m(p)$，如果对所有的 $p \in M$，$m(p)$ 都是常数，就称此 M 为一个 CR 子流形，现在 $H_p(M) = JT_p(M) \cap T_p(M)$ 是 J 不变子空间，由于 $-J^2 = id$，因此 J 作用在 $H_p(M) \otimes C$ 上正好有 $m(p)$ 个以 $+i$ 与 $m(p)$ 个以 $-i$ 为特征根的特征向量，今按此分解 $H_p(M) \otimes C = H_p^{1,0}(M) \oplus H_p^{0,1}(M)$. 如 M 是一个 C^n 中的 CR

子流形，$f:M\to C$ 的一个复值可微分函数，如果对 $\forall V\in H_p^{0,1}(M)$ 有 $Vf=0$，就称 f 是 M 上 CR 函数. 因此由定义可以知道一个 CR 函数亦即满足部分 Cauchy-Riemann 方程的函数. 近来不少数学家对 CR 函数的讨论怀有兴趣，中心问题之一是任一个定义在 $p\in M$ 邻近的 CR 函数，何时可以扩充为 p 点在 C^n 的某个领域的全纯函数. 另一问题是如果可以扩充，那未扩充的范围有多大?这个问题是自然的，因为 C^n 或 C^n 内某个开集上的全纯函数限制在 M 上一定是 M 上的 CR 函数，现在正好是这个自然现象的反问题. 但是这个问题困难程度确实亦是可以预料的，因为一个 CR 函数实际只适合部分 Cauchy-Riemann 方程，而全纯函数是性质极好的函数. 现在关于全纯扩充只有十分初步的结果，讨论主要限制在一般型的 CR 子流形上. 所谓一般型是指 $\mathrm{codim}M=\dim_R M-2m$，这种 CR 子流形每点的法空间，在 J 变换下都映入 $T(M)$ 之中. 宏观看来这个要求是必须的，因为一个 CR 函数要扩充到整个 C^n 上去，主要是解决法向上的扩充，因此每点法空间就必须与原来切空间有紧密的关系. 现在所知的关于 CR 扩充的大部分结果都是集中在超曲面，这时它的法空间是 1 维的，因此超曲面是最自然的一般型 CR 子流形. 现在对 CR 扩充所给出的充分条件是定义一种高阶 Levi 形式 $L_p^l:H_p^{1,0}(M)\to N_p(M)$，$l=1,2,\cdots$，$N_p(M)$ 是法空间，一般的条件是某一次 Levi 形式能满射 $N_p(M)$. 这时局部全纯扩充可以实现，而这种映射或满映射又常常与 CR 流形的局部几何性质有关系. 关于 CR 流形另一个有兴趣的问题是一个抽象的 CR 流形何时可以实现(即局部嵌入)为 C^n 中的一般型 CR 子流形. 对于这个问题 T. Akahari[2] 证明了当 $m\geqslant 4$ 时，拟凸的超曲面型(即 $\dim_R M-2m=1$) 抽象 CR 流形总可以嵌入. 而又有例子证明当 $m=2$ 时一般不成立，唯独 $m=3$ 时这个问题是一个悬案. 现在有不少数学家在关心这个悬案，而关于高余维的抽象 CR 子流形的嵌入问题还几乎未见有什么讨论. 关于 CR 子流形与 CR

扩充是近年来又热闹起来的课题,有兴趣的读者可参考[5,8,9].

§2.6 $\bar{\partial}$问题

设 Ω 是 C^n 的一个域,如 $f = \sum_{\alpha=1}^{n} f_\alpha \mathrm{d}\,\bar{z}^\alpha$ 是 Ω 上的一个(0,1)形式,并满足$\bar{\partial}f = 0$,$\bar{\partial}$问题是指 $\bar{\partial}u = f$ 何时有解,这里$\bar{\partial} = \sum_{\alpha=1}^{n} \mathrm{d}\,\bar{z}_\alpha \frac{\partial}{\partial \bar{z}_\alpha}$ 就是通常的外微分运算.这个问题最成功的例子就是证明Levi 问题,每个拟凸域必是全纯域,解决这个问题的方法就是 L. Hörmander[34] 所提出的 L^2 逼近方法,这个方法是运用泛函与先验估计来求得$\bar{\partial}$问题的解.所得的结果是当 Ω 是拟凸域时,如果 $\int_\Omega |f|^2 e^{-\varphi} < +\infty$,则有解 u 适合 $\int_\Omega |u|^2 e^{-\varphi} \leqslant c \int_\Omega |f|^2 e^{-\varphi}$,特别当 f 是光滑时,u 也是光滑的.由上述结果适当的选取 f 与 φ就可以证明Ω是全纯凸.当 Ω 是强拟凸域与 f 在$\bar{\Omega}$ 上有定义时 G. M. Henkin[33] 给出了$\bar{\partial}$问题解的积分表达式,这一积分表达式提供了对解的一致估计的可能性.因此随着 Henkin 的积分表示公式有不少关于$\bar{\partial}$问题解的一致估计的研究,但是因为直接计算的困难性,这些估计主要是对 C^2 的区域作的.除了一致估计亦有不少人讨论 Henkin 解的 L_p 估计.Levi 问题解还有其他的应用,前面提到的 Siu-Yau 关于单值化定理的证明中就应用了在 Stein 流形上 Levi 问题解的性质.

关于$\bar{\partial}$问题有关的另一个问题是$\bar{\partial}$-Neumann 问题.假定 Ω 是 C^n 或一个 Hermitian 流形上的区域,如果$\bar{\partial}u = f$ 有解,自然此时$\bar{\partial}f = 0$.这个解并不是唯一的,因为对 $\forall\, g \in \mathscr{I}(\Omega)$,$u + g$ 依然是有 $\bar{\partial}(u + g) = f$.因此在$\bar{\partial}u = f$ 解中有一个唯一的解,它与 $\mathscr{I}(\Omega)$ 是正交的(关于通常 $L^2(\Omega)$ 的内积).现在定义这个垂直 $\mathscr{I}(\Omega)$ 的解记之为 Sf,这里 S 表示一个线性算子.现在令 $H^{0,1}(\Omega) = L_2^{0,1}(\Omega)$

$\cap \mathfrak{n}(\bar\partial) \cap \mathfrak{n}(\bar\partial^*)$，这里$\bar\partial^*$表示$\bar\partial$的共轭算子，$\mathfrak{n}(\partial)$与$\mathfrak{n}(\bar\partial^*)$表示对应算子的核空间，所谓$\bar\partial$-Neumann问题是给定$\alpha \in L_2^{0,1}(\Omega)$，$\alpha \perp H^{0,1}(\Omega)$，求$\varphi \in L_2^{0,1}(\Omega) \cap \mathrm{dom}(\bar\partial) \cap \mathrm{dom}(\bar\partial^*)$使其适合$(\bar\partial\bar\partial^* + \bar\partial^*\bar\partial)\varphi = \alpha$，$\varphi \in H^{0,1}(\Omega)$.

当$\bar\partial$-Neumann问题可解时，对每个$\alpha \in L_2^{0,1}(\Omega)$，$\alpha \perp H^{0,1}(\Omega)$，定义$N: L_2(\Omega) \to L_2(\Omega)$；

$$N\alpha = \begin{cases} 0, & \text{如 } \alpha \in H^{0,1}(\Omega); \\ \varphi, & \text{如 } \alpha \perp H^{0,1}(\Omega). \end{cases}$$

则上面定义的$S = \bar\partial^* N$，因此可定义Bergmann投影算子$P: L_2(\Omega) \to \mathscr{I}(\Omega)$如下，对$\forall\, g \in L_2(\Omega)$，

$$Pg = g - S(\bar\partial g) = g - \bar\partial^* N \bar\partial g.$$

如果区域Ω上$\bar\partial$-Neumann问题是可解的，则就知道，在该区域上如果局部L_2Levi问题有解，整体L_2Levi问题就一定有解. 同样在这个区域上当$g \in C^\infty(\overline{\Omega})$，$pg \in C^\infty(\overline{\Omega})$，在前面我们介绍全纯映射时，曾提及过保证区域之间双全纯映射可以扩充到边界的条件R，实际上就是指这个区域Ω，具有当$g \in C^\infty(\overline{\Omega})$，使$pg \in C^\infty(\overline{\Omega})$的性质. S. Bell[6]利用$\bar\partial$-Neumann问题证明$F: \Omega_1 \to \Omega_2$都是拟凸域，如果$\Omega_1$具有条件$R$，则$F$可以扩充到$\overline{\Omega}_1$到$\overline{\Omega}_2$上的$C^\infty$同胚，而且进一步还对全纯逆紧映射的扩充问题作了讨论.

上面提到了多复变函数近年来的某些研究方向，这些并不是多复变函数研究的所有主要内容，有些研究方向同样十分重要，但由于篇幅限制与牵涉的内容较难作比较通俗的介绍，因而只能放弃. 但是有必要在最后提及一下，一是关于紧复流形与紧Riemann曲面的研究，这方面已有极其丰富的成果，其中包括Hodge定理，Riemann-Roch定理，全纯向量丛，陈示性类这些十分重要的内容，有兴趣的读者可以在文献[39,50]中找到满意的介绍. 另外复结

构的形变理论现在亦为大家所重视. 关于全纯函数零点的讨论已是一个专门的分支,这就是解析簇与复空间理论,这方面的理论与代数几何有十分密切的联系. 关于有界域或复流形上的 Bergmann 度量,Kobayashi 度量与 Carathéodory 度量的研究,特别是研究这些度量在边界处的渐近性质,在过去 20 年中亦十分活跃. 关于多复变函数几何理论,多复变函数积分表示理论以及 C^n 中的有界域上函数空间的研究,在过去 20 年均有不少值得称道的成果. 虽然多复变函数近 20 年来有较快的发展,但还有很多基本问题有待彻底解决,而由于多复变函数与其他学科的交错,又不断提出新的有趣的问题需要人们去研究.

参 考 文 献

1. Ahlfors, L. V., Complex Analysis, 3rd ed. McGraw-Hill: New York, 1979.
2. Akahori, T., Contemp. Math., 49(1985), A. M. S. Providence, 1—10.
3. Andreotti, A. and Grauert, H., Bull. Soc. Math. France, 90(1962), 193—259.
4. Arakelyan, N. U., Dokl. Akad. Nauk SSSR, 170(1966), 999—1002.
5. Baouendi, S. and Traves, F., Ann. of Math., 51(1984), 77—107.
6. Bell, S., Ann. of Math., 14(1981), 103—113.
7. Bell, S. and Ligosga, E., Invention Math., 57(1980), 283—289.
8. Boggess, A., Duke Math. J., 48(1981), 665—684.
9. Boggess, A. and Pitts, J., Duke Math. J., 52(1985), 67—102.
10. de Brange, L., Acta Math., 154(1985), 137—152.
11. Burns, D., Shnider, S. and Wells, R., Invention Math., 46(1978), 237—253.
12. Carleson, L., Amer. J. Math., 80(1958), 921—930.
13. Carlson, J. and Griffiths, P., Ann. of Math., 95(1972), 557—584.
14. Chern, S. S., Ann. of Math., 71(1960), 345—365.

15. Chern, S. S. and Moser, J., Acta Math., 133(1974), 219—271.

16. Chuang, C. T. and Yang, L., Contemp., Math., A. M. S. 48, 1985, 21—64.

17. Clunie, J., J. London Math. Soc., 42(1967), 389—392.

18. Collingwood, E. F. and Lohwater, A. J., The theory of cluster sets, Cambridge Univ. Press, 1966.

19. Drasin, D., Acta Math., 128(1977), 81—151.

20. Drasin, D., Acta Math., 158(1987), 1—94.

21. Duren, P., Theory of H^p spaces, Academic Press, New York, 1970.

22. Duren, P., Univalent functions, Springer-Verlag, 1983.

23. Edrei, A. and Fuchs, W., Comment. Math. Helv., 33(1959), 258—295.

24. Eremenko, A. E., Sibersky Math. J., 27(1986), 87—102.

25. Fefferman, C., Invention Math., 26(1974), 1—65.

26. Fuchs, W., Ann. of Math., 68(1958), 203—209.

27. Grauert, H. and Remmert, R., Math. Z., 65(1956), 175—194.

28. Greene, R. E. and Wu, H., Lecture Notes in Math., No 699.

29. Greene, R. E. and Wu, H., Acta Math., 137(1976)209—245.

30. Hayman, W. K., Meromorphic functions, Oxford, 1964.

31. Hayman, W. K., Research problems in function theory, Athlone Press, 1967.

32. Hayman, W. K. and Ynag, L., Proc. London Math. Soc., 44(1982), 193—214.

33. Henkin, G. H., Math. Sb., 78(1969), 611—632.

34. Hörmander, L., An introduction to complex analysis in several complex variables, North-Holland Pub. Co., 1973.

35. 华罗庚,多复变函数论中典型域的调和分析,科学出版社,1958.

36. Lewis, J. L. and Wu, J. M., J. d'Analyse Mathématique, 50(1988), 259—283.

37. Makarov, N. G., Proc. London Math. Soc., (3)51(1985), 369—384.

38. Mori, S., Ann. Math., 110(1979), 593—606.

39. Morrow, J. and Kodaira, K., Complex manifolds, Holt, Rinehart and

Winston Inc. ,1971.

40. Narasimhan, R. , Several complex variables, The Univ. , of Chicago Press, 1971.
41. Nevanlinna, R. , Analytic functions, Springer-Verlag, 1970.
42. Osgood, C. , J. of Number Theory, 21(1985), 347－389.
43. Pfluger, A. , Comment. Math. Helv. , 19(1946), 91－104.
44. Schwick, W. , J. d'Analyse Mathématique, 52(1989), 241－290.
45. Siu, Y. T. and Yau, S. T. , Ann. Math. , 105(1977), 225－264.
46. Siu, Y. T and Yau S. T. , Invention Math. , 59(1980), 189－204.
47. Steinmetz, N. , J. Reiene Angew. Math. , 368(1986), 134－141.
48. Stoll, W. , Bull. of A. M. S. , 32(1977), 166－183.
49. Weitsman, A. , Acta Math. , 128(1972), 41－52.
50. 伍鸿熙，吕以辇，陈志华，紧黎曼曲面，科学出版社，1981.
51. Yang, L. , J. London Math. Soc. , 25(1982), 288－296.
52. Yang, L. , Indiana Univ. Math. J. , 35(1986), 179－192.
53. Yang, L. , Trans. Amer. Math. Soc. , 308(1988), 583－601.
54. Yang, L. , J. d'Analyse Math. , 55(1990), 287－296.

Ⅶ

代数几何与代数数论

Algebraic Geometry and Algebraic Number Theory

肖　刚
（华东师范大学数学系）
Xiao Gang
(Department of Mathematics,
East China Normal University)

低维代数流形：历史与发展

Algebraic Manifolds with Lower-Dimensions: History and Progress

中国数学发展的若干主攻方向
Some Main Aspects on Development
of Mathematics in China

代数几何的发展历史经历了一个有趣的反复：从具体到抽象，又回到具体.

一般认为，代数几何的研究是从 Abel 关于椭圆积分的研究开始的(椭圆曲线). 随后 Riemann 引入了曲线的亏格(genus) 概念并获得了一般亏格的曲线的一系列基本性质，甚至提出了曲线的模空间概念，从而完成了代数曲线研究的基本框架. 19 世纪和 20 世纪初，Max Noether 和以 Castelnuovo，Enriques，Severi 等人为代表的意大利学派完成了一个十分漂亮的代数曲面分类理论. 至此为止，代数几何研究的内容主要局限于曲线和曲面，而且不是建立在一个十分严格的理论基础上的，以至于意大利学派的一些从直观出发"证明"的结论以后被发现是错误的. 然而在 20 世纪的 50 到 60 年代，以 Serre 和 Grothendieck 为首的法国学派通过层的上同调和概型的概念，不仅建立了任意域甚至任意环上的统一且严谨的代数几何理论，而且该理论甚至可以用来描述像代数数论这样的相近学科. 而代数几何研究的中心兴趣也从层、概型、étal 上同调到结晶上同调，迅速上升到抽象的顶端. 但随着 60 年代末 Grothendieck 的隐退，该"抽象热"渐渐降温；而同时，复数域上低维流形的研究又由于一系列重要成果的建立而吸引了越来越多的注意，从而重新占据了代数几何研究的主要位置：如 Mumford 和 Harris 等人建立的曲线的模空间理论，Bombieri 关于曲面的多重典范映射的结论和宫冈洋一-丘成桐不等式，以及最近Donaldson-Freedman 的四维拓扑流形理论和森重文等人的曲体(3-fold) 理论等.

限于篇幅及作者的兴趣，我们只介绍代数几何研究中最具体的那部分：复数域上低维的射影代数簇(曲线、曲面和曲体)，并着重于它们的分类问题.

§1 什么是代数几何

设 P 是复数域 C 上的一个 n 维射影空间. 我们可以给 P 确定一组齐次坐标 $x_0,\cdots,x_n$. 关于 $x_0,\cdots,x_n$ 的一个多项式 $F(x_0,\cdots,x_n)$ 称为 d 次齐次多项式,如果 F 的每一项的总次数都是 d. 在这种情况下,使 F 为零的那些 $x_0,\cdots,x_n$ 就构成 P 中的一个几何流形. P 中对应于这样的 $x_0,\cdots,x_n$ 的一个点也叫做 F 的一个零点.

一般地,我们可以考虑 P 中由一组(有限或无限多个,不一定同次)齐次多项式的公共零点所定义的流形. 这样的一个流形叫做 P 中的(射影)代数集. 因为两个代数集的并也是一个代数集,在很多情形下人们只需要考虑不可约的代数集,也就是不能表示成两个真子代数集之并的代数集. 这样的代数集就叫做 P 中的射影代数簇,它是代数几何研究的基本对象.

设 V 是一个代数簇. 代数几何感兴趣的不仅是 V 作为拓扑流形的几何特性,而且更多的是 V 上的代数结构. 这个代数结构是由 V 上的所有有理函数(或称代数函数)所确定的. 每个有理函数 f 都可以表示成两个相同次数的齐次多项式的商,并且我们要求分母不在 V 上恒等于零,于是 f 在 V 的一个处处稠密的开子集上有定义. 这是一种代数几何特有的开子集,即所谓 Zariski 开子集.

V 上的所有有理函数自然地构成一个域 $K(V)$,称为 V 的有理函数域. 我们可以定义 V 的维数为 $K(V)$ 在复数域 C 上的超越次数,事实上这样定义的维数就等于 V 作为复数域上解析流形时的复维数. 1 维的代数簇又称代数曲线,2 维的称为代数曲面,3 维的称为曲体.

另一方面,代数几何关心的首先是代数簇 V 上的代数结构而不是 V 在空间 P 中的嵌入. 在这个意义下,如果我们有两个代数簇

V 和 W 之间一个一一映射 $\varphi: V \to W$，它把 W 上的有理函数对应成 V 上的有理函数且反之亦然，则 φ 被看成是 V 和 W 之间的一个同构映射并且 V 和 W 因为同构而被认为是代表了同一个代数簇. 例如，射影直线 P^1 和 P^2 中由方程 $x_0^2 + x_1^2 + x_2^2 = 0$ 所定义的二次曲线就是同构的. 于是代数几何学家们往往不把同构于 P^1 的代数曲线称为"直线"而是冠之以一个新的名称：有理曲线，因为它不总是"直"的.

更一般地，我们有两个代数簇 V，W 之间的态射概念：一个映射 $\varphi: V \to W$ 称为态射(morphism)，如果它诱导 W 和 V 的代数结构之间的一个同态，也就是说对于 W 上的每个有理函数 f，复合映射 $\varphi \circ f$ 是 V 上的有理函数. 态射是代数几何中最基本的映射概念，但很多时候它的条件显得太强，所以我们有更一般的有理映射概念：对于有理映射 $\varphi: V \to W$，我们只要求 φ 是 V 的一个 Zariski 开子集到 W 的映射，但当然仍要求 $\varphi \circ f$ 是 V 上的有理函数. 特别地，如果有理映射 φ 有一个有理逆映射，我们称 V 和 W 是双有理等价的. 这时虽然 V 和 W 不一定同构，但它们的区别其实很小，比如说有理函数域 $K(V)$ 和 $K(W)$ 就是同构的.

双有理等价概念对于代数簇的分类问题有着关键的意义：一般来说，任一代数簇都有无限多个与其双有理等价但不同构的代数簇. 但因为双有理等价的代数簇在整体上有相当重要的共同性质而它们的不同只是局部的，可以很自然地把这样的代数簇看成是同一类的. 所以通常所说的代数簇的分类实际上是对代数簇的双有理等价类的分类. 此外，根据著名的广中平佑奇点解消定理，任意代数簇都双有理等价于一个光滑代数簇(或称非奇异代数簇)，即没有奇点的代数簇. 这样至少在理论上，对代数簇的双有理等价类的分类及其整体性质的研究就可以化为对光滑代数簇的双有理等价类的这样的研究，从而避免了局部的奇异点的存在对整体性质研究可能带来的干扰. 在曲线和曲面的情形，这样的考虑确

实是很有效的，虽然我们下面可以看到，从三维情形开始，仅考虑光滑簇是不够的，必须同时允许一些特殊的奇异点的存在才能克服由于没有合理的极小模型而带来的困难.

如果说代数簇 V 上有理函数定义了 V 的基本代数结构的话，对研究 V 的整体性质并对其进行分类的最重要的工具是 V 上的层(sheaf). 其中人们研究得最多的是局部自由层，就是由 V 上的某个解析向量丛的所有局部截面(local section) 所构成的层. 这里 V 上的一个解析向量丛是一个解析空间 M(可以理解为带奇点的解析流形)，以及从 M 到 V 的一个解析映射 $\varphi: M \to V$，使得对于 V 的每个点 P，$\varphi^{-1}(P)$ 是一个具有固定维数 r 的复向量空间. 数 r 就叫做 M 的(或者对应的局部自由层的) 秩. 事实上，取局部截面的过程构成了 V 上的所有解析向量丛和所有局部自由层所成的集合之间的一个一一对应，所以有时人们往往不加区别地混用向量丛和局部自由层的概念. 当 V 为光滑代数簇时，向量丛的一个明显的例子是 V 上的切空间构成的秩为 $\dim(V)$ 的切丛，对应于 V 的切层 T_V. 这时 V 上的所有一阶微分形式也自然构成一个秩为 $\dim(V)$ 的局部自由层 Ω_V.

当局部自由层 L 的秩为 1 时，L 称为可逆层，相应的向量丛称为线丛. V 上所有可逆层的全体以张量积为运算自然地形成一个群，这就是 V 的 Picard 群 PicV，其单位元对应的是 V 上所有有理函数构成的可逆层，称为平凡层.

可逆层是最常见也是最有用的层，因为它们与代数簇到射影空间中的映射有着密切的关系：

设 L 为代数簇 V 上的一个可逆层. L 中的所有整体截面(在 V 上处处有定义的截面) 构成复数域上的一个有限维的向量空间，记作 $\Gamma(L)$ 或 $H^0(L)$，它的维数记作 $h^0(L)$. 当 $\Gamma(L)$ 非空时，这些整体截面自然地定义了 V 到射影空间中的一个有理映射 $\varphi_L: V \to P^n$，这里 $n = h^0(L) - 1$. 特别当 φ_L 为嵌入映射时，L 称为非常丰富

层(very ample). 反之,若 $\varphi: V \to P^n$ 是一个嵌入态射,则 P^n 中的一次齐次形式自然地诱导 V 上的一个非常丰富层 L,使得 $\varphi = \varphi_L$. 所以 V 上的非常丰富层一一对应于 V 在射影空间中的表示.

假设 V 是一个维数为 d 的光滑代数簇,则 V 上所有的局部 d 阶外微分形式构成一个可逆层 ω_V,叫做 V 的典范层,它所诱导的有理映射称为典范映射. 我们有 $\omega_V = \Lambda^d \Omega_V$. 一般地,对于任一正整数 n,我们有 n- 典范层 $\omega_V^{\otimes n}$ 及其对应的 n- 典范映射. $h^0(\omega_V^{\otimes n})$ 记作 $p_n(V)$,或 p_n. 而当 $n=1$ 时,p_1 又可记作 p_g,称为 V 的几何亏格. 所有这些 n- 典范层以及 n- 典范映射在 V 的同构意义下都是唯一确定的,因此是 V 上重要的几何对象. 不仅如此,所有的 p_n 都是双有理不变量,它们对于双有理等价的代数簇是不变的. 于是下面的典范模型也是双有理不变的:

设 $R = \bigoplus \Gamma(\omega_V^{\otimes n})$,这里求和是对所有大于或等于 1 的 n 作的. 我们假设 R 不是空集,则 R 是 C 上的一个无限维向量空间,而且 n- 典范层之间的张量积关系在 R 上诱导了一个分次环结构,称为 V 的典范环. 在低维的情形,人们已证明 R 是有限生成的. 这时 R 可以自然地定义一个射影代数簇 Proj(R),这就是 V 的典范模型,它的维数不超过 $\dim(V)$.

我们向读者推荐[15] 作为一本很好的代数几何入门书. 当然,最好的正规教科书仍然是[9].

§2 曲线:高维情形的缩影

在一维的情况,代数几何的很多结论都变得非常简单. 例如,对代数曲线的双有理等价类的研究与对光滑代数曲线的研究事实上是一回事,因为每个这样的双有理等价类中都唯一地存在一条光滑曲线. 这使我们可以只考虑光滑曲线的分类和整体性质的研究,因此我们以下所指的曲线都是光滑的射影代数曲线. 一条这样

的曲线 C 作为微分流形就是一个可定向的紧致 Riemann 面，因而拓扑同胚于一个有 g 个眼的环面：

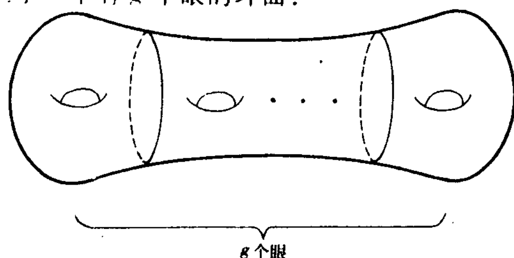


图 1　有 g 个眼的环面

数 g 是由 C 唯一确定的，叫作 C 的亏格，记为 $g(C)$. 亏格也可以用代数的方法来定义，因为 $g(C) = h^0(\omega_c)$. 对每个非负整数 g，都有一条代数曲线 C 使得 $g(C) = g$. 而亏格不同的曲线显然是不同构的，所以亏格是曲线的一种"数值不变量". 而曲线在一个特定的射影空间中的嵌入下的次数不是数值不变量，如 1 次的直线可以同构于一条 2 次的曲线. 正因为亏格"不变"，它给曲线的分类提供了一个重要的基础.

亏格为零的曲线只有一条，即有理曲线. 而亏格为 1 的曲线又称椭圆曲线，这是一个代数群：这条曲线同时又是一个群，并且群运算所诱导的映射都是代数几何意义下的态射. 椭圆曲线作为群来说可能并不十分有趣，因为它们都是很简单的交换群. 但作为代数曲线却一直吸引着人们浓厚的兴趣. 椭圆曲线的分类有几种不同的途径，其中之一是通过椭圆曲线在射影平面中的嵌入和 j- 不变量：每条椭圆曲线 C 都可以表示成射影平面中的一条光滑三次曲线，这时有一个三次方程作为它的定义方程. 通过合适的坐标变换，这个定义方程可以写成形式 $x_0x_2^2 = x_1(x_1 - x_0)(x_1 - \lambda x_0)$，这里 λ 是一个复数，而 C 的 j- 不变量就定义为

$$j = 2^8 \frac{(\lambda^2 - \lambda + 1)^3}{\lambda^2(\lambda - 1)^2},$$

它不依赖于C在平面中的具体嵌入和坐标的选取，所以称为“不变量”. 另一方面，对每个复数α，一定存在唯一的一条椭圆曲线C，使得α等于C的j-不变量. 这样，j-不变量就给出了所有椭圆曲线的集合$\mathscr{M}_1$与复数域上的仿射直线A(注意A是射影直线P^1中的一个Zariski开子集)中的点的一个一一对应. 不仅如此，这个对应是个很自然的代数对应：P^1上的代数结构诱导A上的一个代数结构，而A中的点在这个代数结构下的“形变”与它所对应的椭圆曲线在代数意义下的自然形变是一致的.

椭圆曲线的这个分类理论给我们提供了一个有益的启发：设$\mathscr{M}_g$为所有亏格为g的曲线所成的集合，这里设$g \geqslant 2$. 是不是可以在$\mathscr{M}_g$上赋予一个代数结构，它与曲线在代数意义下的自然形变相一致，使得带有这个代数结构的$\mathscr{M}_g$成为某个射影代数簇的一个Zariski开子集(这样的流形叫做拟射影代数簇)？如果这样的代数结构存在，则$\mathscr{M}_g$称为亏格g的曲线的一个模空间(moduli space).

Riemann 在研究代数曲线(即 Riemann 面)的时候就已经有了模空间的想法，有趣的是，虽然他未能证明这样一个代数结构的存在性，却发现$\mathscr{M}_g$的维数一定是$3g-3$. 模空间的严格理论是Mumford 在 60 年代通过把古典的不变量理论发展成几何不变量理论，从而建立起来的. Mumford 的办法是，因为 3-典范丛是非常丰富的，通过3-典范映射，每条亏格$g \geqslant 2$的曲线都可以表示成$5g-6$维空间P中的一条$6g-6$次曲线，因此满足P中的一个Hilbert 多项式. 而P中满足这个 Hilbert 多项式的所有子代数簇的集合本身构成一个代数簇H. Mumford 引入了稳定曲线的概念，使得所有亏格为g的稳定曲线在 3-典范映射下的象的集合成为H中的一个子簇M. P上所有射影变换所成的一般射影群G自然地作用于M上. Mumford 证明M上所有在G的作用下不变的代数形式所成的环对应于一个射影簇，它正好是M在G的作用下的商，因此就是含$\mathscr{M}_g$为处处稠密开子集的一个射影代数簇.

今天，人们对 $\mathscr{M}_g$ 的性质已经有了相当深刻的了解[8].

§3 曲面：从意大利学派发展而来

从曲线到曲面的第一个困难是，双有理等价类中的光滑曲面不再是唯一的. 相反，每个曲面都双有理等价于无限多个光滑曲面. 因此，曲面的分类要做的第一件事就是如何在曲面的每个双有理等价类中找到一个可以唯一确定的合理的"模型". 虽然这种模型不是对所有的双有理等价类都存在的，至少对于绝大多数曲面来说，极小曲面可以是这样的一个模型.

首先使人们想到的最显然的曲面是两条曲线的积：$S = C_1 \times C_2$. 如果其中的一条曲线，比如说 C_1，是有理曲线的话，曲面 S(或双有理等价于 S 的曲面) 就叫作直纹面，它是十分特殊的一类曲面. 例如射影平面 P^2 和 n 维空间中次数低于 $2n-2$ 的曲面都是直纹面.

意大利学派关于曲面分类的一个重要定理就是，除了直纹面之外，每个代数曲面都双有理等价于唯一的一个"极小曲面". 极小曲面是这样的光滑曲面，任何从它到另一个光滑曲面的双有理态射都是同构. 另一个等价的定义是：极小曲面是不包含自相交数为 -1 的有理曲线(称为例外曲线) 的光滑曲面，因为如果一个曲面 S 包含一条例外曲线 C，则存在 S 到另一个光滑曲面 S' 的一个双有理映射 $f: S \to S'$，它把 C 映到 S' 中的一个点(即 f 收缩 C)，而在 C 以外是同构. 从任何曲面开始，最多只能做有限多次这样的收缩，所以任何直纹面也双有理等价于某个极小曲面，但这样的极小曲面不唯一，例如 $P^1 \times P^2$ 和 P^2 是双有理等价的，但不同构.

极小曲面的存在，使得除了人们对其性质相当清楚的直纹面之外，对代数曲面的双有理等价类的分类问题就转化成了对极小曲面的分类问题，使得这个问题有了十分容易处理的模型，并且使

我们可以考虑这样的双有理等价类所因此而自然地对应着的一些几何对象，其中最重要的莫如极小模型的陈类 c_1 和 c_2，以及由此产生的两个陈数 c_1^2 和 c_2(这里实际指 c_2 的次数). 陈数是曲面的数值不变量，它们对于曲面分类的意义就如同亏格在曲线的情形. 所以，我们以下所称的曲面都是指极小曲面.

除了直纹面外，还有一些曲面由于其特殊性而从一开始就引起人们的很大兴趣并得到了深入的研究：

1. 椭圆曲面. 如果曲面 S 有一个到某一条曲线 C 的有理映射 $f:S \to C$，使得对于 C 上几乎所有的点 P，P 在 f 下的原象是 S 中的一条椭圆曲线，S 就称为椭圆曲面.

2. $K3$ 曲面. $K3$ 曲面的严格定义是第一陈类 c_1 等于零的单连通曲面. 作为一个众所周知的例子，3 维射影空间中由方程 $x_0^4 + x_1^4 + x_2^4 + x_3^4 = 0$ 所定义的 4 次曲面就是一个 $K3$ 曲面. $K3$ 曲面的陈数满足 $c_1^2 = 0, c_2 = 24$.

3. Abel 曲面. 它的定义就是同时是代数群的代数曲面，因而在某种意义上来说可以认为是椭圆曲线在二维情形的一种自然的推广. 这些曲面的陈数满足 $c_1^2 = c_2 = 0$.

与上面这些特殊的曲面相对的是一般型曲面. 顾名思义，有时可以说几乎所有的曲面都是一般型曲面.

如果一个曲面 S 的典范模型也是一个曲面，则 S 称为一般型曲面. 一般地，一个代数簇 V 的典范模型的维数称为 V 的小平(Kodaira) 维数，记作 $\kappa(V)$. 当典范模型为空集时，$\kappa(V)$ 定义为 $-\infty$. 若 $\kappa(V) = \dim(V)$，则 V 称为一般型的. 小平维数是一个很重要的双有理不变量，在代数簇的分类中起着关键的作用. 在一维的情形，曲线 C 是一般型的当且仅当 $g(C) \geqslant 2$. 而曲面 S 的小平维数为 $-\infty$ 当且仅当 S 为直纹面；$K3$ 曲面和 Abel 曲面的小平维数都等于零，而椭圆曲面的小平维数可以是 $-\infty$，0 或者 1. 所以这些曲面都不是一般型的.

另一方面，可以方便地举出很多一般型曲面的例子，如两条曲线的积 $S = C_1 \times C_2$，当因子 C_1 和 C_2 都是一般型曲线时，S 是一般型曲面. 3维射影空间 P^3 中次数超过4的光滑曲面也是一般型的.

曲面分类理论的中心定理，简单地说就是：任一代数曲面 S 必为上面所定义的五种曲面（直纹面，椭圆曲面，$K3$ 曲面，阿贝尔曲面，一般型曲面）之一.

意大利学派的这个曲面分类定理尽管十分漂亮，却也不是十全十美，因为它对几乎所有曲面（一般型曲面）的分类没有给出任何信息. 所以进入现代以来，人们对曲面分类问题的注意力就集中到了一般型曲面上来.

现代的曲面研究热是从Bombieri在1973年的著名论文[3]开始的. 简单地说，Bombieri 证明了当 $n \geqslant 5$ 时，一般型曲面的 n-典范映射是双有理态射. 这就给 Gieseker 随后利用 Mumford 的几何不变量理论证明对每一对固定的陈数 (c_1^2, c_2)，一般型曲面的模空间是有限多个拟射影簇的并[6]提供了基础. 然而与曲线的情形不同，对于这些拟射影簇的维数和个数，目前已知的基本上只有 Catanese 关于维数的一个不很精确的估计[4].

需要注意的是，Gieseker 定理并没有说明在陈数 (c_1^2, c_2) 给定的情形下，一般型曲面是不是存在. 事实上，从经典的结果人们知道，极小一般型曲面的陈数必须满足 $c_1^2 + c_2 \equiv 0 \pmod{12}$，$5c_1^2 \geqslant c_2 - 36$，$c_1^2 > 0$，$c_2 > 0$. 在1976年，宫冈洋一和丘成桐同时证明了另一个重要的不等式：$c_1^2 \leqslant 3c_2$. 因为存在无限多个分别使上面两个不等式中等号成立的陈数对 (c_1^2, c_2) 以及对应的曲面，这些不等式是不能再改进的. 于是现在一般型曲面的分类中的一个重大问题就是，对于满足上述所有条件的 (c_1^2, c_2)，是不是能找到一个对应的极小一般型曲面. 这个问题被称为曲面的地理问题，当 $c_1^2 \leqslant 2c_2$ 时，Persson 在1981年对此作了肯定的回答[11]；基于肖刚的一个例子，陈志杰又把 Persson 的条件减弱为 $c_1^2 \leqslant 2.7c_2 - c$，其中 c 是一

个常数[5],[16].根据肖刚最近的一个改进的方法,条件中c_2的系数可以进一步增加到2.84(尚未发表),这是目前最好的结果.

即使曲面的地理问题在陈数的意义下得到了完整的解决,还有更困难的一面:除了陈数之外,代数曲面还有别的双有理数值不变量.例如几何亏格p_g,它与陈数并不是完全相关的.包含p_g在内的曲面地理问题研究则尚处于比较原始的状态,目前只知道一些不完整的不等式.

至于一般型曲面的n-典范映射,虽然Bombieri已经解决了绝大多数的情形,但在$n = 2$时(双典范映射)他未把所有的情形都算出来,特别是Bombieri完全没有考虑$n = 1$(典范映射)的情形.曲面的双典范映射方面,由于最近Igor Reider, Francia和肖刚的一系列工作,有了很大进展,未能解决的情形已经很少(参见[18]);而典范映射与曲线情形和曲面的多重典范映射不同,情形变得极为复杂.这方面的第一个工作是Beauville的[2],其中的一部分不等式随后被肖刚改进成了最佳形式(参见综合报告[4]),但目前为止典范映射方面还有很多有待解决的问题.

一般型曲面的几何性质也有不少棘手的难题,很多在曲线情形很容易证明的结论,对于曲面变得非常困难.例如,亏格为$g \geqslant 2$的曲线的自同构群是不超过$84g - 84$阶的有限群,这是一个经典的定理.虽然人们知道一般型曲面的自同构群也是有限的,但对其阶的上界估计却只有一些很不完整的结果,如该上界不超过c_1^2的一个多项式函数.若G是这个自同构群的一个交换子群,则G的阶不超过$52c_1^2 + C$,其中C是一个常数[17].(1991 — 1992年,肖刚证明了一般型极小曲面的自同构群的阶不超过$42^2{c_1}^2$,并且这是最好的界[19]).

总而言之,代数曲面的研究发展到今天,已经形成了一个十分庞大的理论,其中包括了很多漂亮的定理,但是对比曲线理论的完整性,似乎尚待解决的问题比已经取得的成果还是多得多.关于曲

面研究的最新进展的更详细的综述，读者可以参看[4]，[12].

§4　曲体：崭新而艰难的理论

曲体就是三维的代数簇.有趣的是，从曲面到曲体的困难一点不比从曲线到曲面所遇到的少.第一个困难是，在三维的情形，一般的双有理等价类中都不存在合理的光滑极小模型.正是这个困难长期以来一直阻碍了曲体分类研究的所有努力.

突破是从80年代初森重文的锥理论开始的.森深入研究了一个曲体 X 的所有有理一维链(1-cycle)的空间 N_1 中，由有效链所张成的锥 C.他发现，这个锥在与 X 的典范除子相交为负的那一部分是一个局部有限的多面体，这个多面体的棱称为极射线(extremal ray) R.于是有森的收缩定理：

X 中对应于一条极射线 R 的曲线是可收缩的.也就是说，存在一个满态射 $\varphi_R: X \to Y$，把对应于 R 的每条曲线都映到 Y 中的一个点.并且如果 Y 的维数也是3，则 φ_R 是双有理的.

如果 Y 的维数小于3，则 φ_R 给出了一个对维数归纳的途径，并且对应于 $\dim(Y)$ 为0，1或2的情形，X 分别为 Q-Fano 曲体，以 del Pezzo 曲面为纤维的纤维空间，或圆锥丛.这些都是小平维数为 $-\infty$ 的曲体.因此分类问题在这种情形下就完成了.

在 φ_R 为双有理的情形下，问题的关键是要证明上述森收缩一定可以终止于某一步.这个收缩到最后的曲体就是我们所要的极小模型，它的几何特征是森锥 C 中没有负部分，换言之 X 的典范层是数值有效(nef)的.

需要指出的是，即使 X 是光滑曲体，Y 也不一定是光滑的，但 φ_R 所带来的奇点都是一类性质较好的奇点，叫做典范奇点.所以在一般情形下，曲体的极小模型即使存在，也不是光滑的，而是带典范奇点的奇异簇.

森收缩的终止问题是森理论中最困难的部分．首先，双有理的森收缩有两种可能：φ_R 或是曲面型收缩，或是曲线型收缩．后者光靠森收缩不能完全解决问题，因为曲线型收缩所导致的 Y 可能含有非常坏的奇点．

最后解决曲体极小模型问题的是森重文 1988 年的翻转定理(flip theorem)[10]：

设 $\varphi_R:X\to Y$ 是一个单纯的曲线型森收缩，则存在另一个单纯曲线型森收缩 $\varphi':X'\to Y'$，使得 Y' 只有典范奇点．

从 X 到 X' 的这种转换过程称为一个正向翻转(directed flip)．苏联数学家 V. V. Shokurov 已于 1985 年证明，从一个给定的曲体 X 出发，最多只能做有限多次正向翻转．因此，如果 X 的小平维数不是 $-\infty$，则一定可以通过有限次的曲面型森收缩和正向翻转，达到一个双有理等价于 X 且不再有森收缩的曲体 X'，即 X 的极小模型．这样就把曲面的极小模型理论成功地推广到了三维的情形．

从森翻转定理还可以立即得出关于曲体的一个重要推论：光滑曲体的典范环是有限生成的，因而典范模型是一个射影代数簇．

由此可以看出，目前人们对于曲体的了解大致上相当于意大利学派在曲面情形所达到的程度．关于曲体研究的更详细的介绍，请参见[6]，[13]，[14]．

参　考　文　献

1. Beauville, Arnaud: Surfaces Algebriques Complexes, Asterisque 54, Soc. Math. France, 1978.
2. Beauville, Arnaud: L'applications canoniques pour les surfaces de type general, Inv. Math. 55, 121－140(1979).
3. Bombieri, Enrico: Canonical models of surfaces of general type,

Publication IHES 42,171—219(1973).

4. Catanese, Fabricio: Canonical rings and "special" surfaces of general type, in Algebraic Geometry: Bowdoin 1985, vol I, 175—194, American Math. Soc. 1987.

5. Chen, Zhijie(陈志杰): On the geography of surfaces——simply connected minimal surfaces with positive index, Math. Annalen 277, 141—164(1987).

6. Gieseker, David: Global moduli for surfaces of general type, Inv. Math. 43, 233—282(1977).

7. Griffiths, Phillip: 代数曲线,北京大学出版社,1985.

8. Harris, Joe: Curves and their moduli, in Algebraic Geometry: Bowdoin 1985, vol I, 99—144, American Math. Soc. 1987.

9. Hartshorne, Robin: Algebraic Geometry, Graduate Text in Math. 52, Springer-Verlag 1977.

10. Mori, Shigefumi(森重文): Flip theorem and the existence of minimal models for 3-folds, Journal of the AMS 1/1, 117—253 (1988).

11. Persson, Ulf: Chern invariants of surfaces of general type, Compositio Math. 43/1, 3—58 (1981).

12. Persson, Ulf: An introduction to the geography of surfaces of general type, in Algebraic Geometry: Bowdoin 1985, vol I, 195—218, American Math. Soc. 1987.

13. Reid, Miles: Young person's guide to canonical singularities, in Algebraic Geometry: Bowdoin 1985, vol I, 345 — 414, American Math. Soc. 1987.

14. Reid, Miles: Tendencious survey of 3-folds, in Algebraic Geometry: Bowdoin 1985, vol I, 333—344, American Math. Soc. 1987.

15. Reid, Miles: Undergraduate Algebraic Geometry, Cambridge Univ. Press, 1988.

16. Xiao, Gang(肖刚): An example of hyperelliptic surfaces with positive index, 东北数学 2(3), 255—257 (1986).

17. Xiao, Gang(肖刚): On abelian automorphism groups of surfaces of

general type, Invent. Math. 102, 619—631(1990).

18. Xiao, Gang(肖刚): Degree of the bicanonical map of a surface of general type, American J. Math. 112, 713—736(1990).

19. Xiao, Gang (肖刚): Bound of automorphisms of surfaces of general type, I, II, Ann. Math.

冯克勤
（中国科学技术大学数学系）
Feng Keqin
（Department of Mathematics，
University of Science and Technology of China）

从 Gauss 到 Langlands——介绍代数数论

From Gauss to Langlands ——An Introduction to Algebraic Number Theory

中国数学发展的若干主攻方向
Some Main Aspects on Development of Mathematics in China

§1 引　言

> 直到最近，数论(或者有人更愿意称它为算术)的研究经久不衰且大放异彩，这不仅在于献身于它的人数众多，更是由于它的精深和优美. 与此同时，或许无与伦比的是它所激发出来的那种热情，那种由 Euler、Gauss、Eisenstein 和 Hilbert 这样一些伟人用各自感人的言辞所表达的狂热之情.
>
> ——André Weil，《从历史看数论》，1983，前言

代数数论是数论的一个分支，它起源于初等数论. 大家知道，初等数论的基本问题是研究整数的算术性质(整除性质、同余性质、素数分布等)以及方程或方程组的整数解，或者说，初等数论是研究有理数域 Q 和它的整数环 Z 的算术特性. 而代数数论，就其古典意义下，则是研究任意代数数域(即有理数域的有限次扩域)和它的(代数)整数环的算术和代数特性. 从 18 世纪直到 20 世纪初，人们在研究数论问题时创造了许多重要的代数概念，将代数学推进到"抽象"代数(或叫近世代数)阶段. 将这些新的代数工具用于数论研究，便建立了古典的代数数论. 无论在历史上还是在当今世界，均有为数众多的杰出数学家致力于代数数论的研究，引入新鲜的和深刻的思想和工具，使这门学科不断地取得重大成果，也不断地提出一些挑战性的新课题和新猜想，吸引有志的年轻人加入代数数论研究队伍的行列. 近年来. 代数数论以其丰富的成果和深刻的思想方法与代数几何、群表示理论、调和分析、微分几何和偏微分方程等许多其他数学领域相互渗透和交织，不仅促进了这些学科的发展，产生了许多边缘性数学分支，而且也大大改变了数论的面貌，形成了现代代数数论的各种分支和流派.

人们常常把数学的抽象性与联系实际对立起来，但实际上，数

学的抽象性恰好决定了它应用的普遍性和广泛性. 数学的深刻性恰好决定了它应用的有效性. 代数数论这门学科是多年来人类智慧和思维的高度创造性的结晶,在国民经济和国防建设中已经发挥了巨大的威力. 1983 年,由丘成桐等美国 11 位著名数学家撰写了一份报告,评述数学研究的最新进展,从以下摘录的其中三段可见代数数论在计算机科学,通信工程等实际领域的重大影响(全文见《数学译林》1985 年第 4 卷第 2 期):

> "计算机的出现使得计算理论成为数学研究的一个领域,它坚实地建立在诸如概率论、组合学、代数几何和数论这些现代数学主流的方法和土壤之上,从而为实际计算机科学家创造了许多重要的工具."
>
> "有效的算法常常有重要的实际价值,值得注意的例子有:快速 Fourier 变换和它在信号处理中的应用,在数论和有限域理论中近来发展的随机化算法和它们在纠错码和密码学中的应用."
>
> "编码理论与密码学的发展是'纯粹'数学在应用领域得到出人意料应用的一个突出的例子. 几年前,A. Weil 于 1948 年的数论工作已经用于编码理论. 去年,一个苏联数学家小组把 Deligne, Rapoport, Ihara 和 Langlands 近来在代数几何这一最抽象领域的工作用于设计达到理论上有效性的纠错码,而设计这种纠错码在以前被认为是不可能的."

限于本文的篇幅和作者的学识水平,这里无法全面介绍代数数论的发展历史,它的全部光辉思想、丰硕的成果以及广泛的应用前景. 我们只试图简要和通俗地介绍代数数论的最基本内容、思想、方法、结果和历史发展的轮廓.

§2 Gauss 和 Kummer——古典代数数论的创始人

从公元前3世纪的Diophantus到17世纪法国的Fermat，对于方程整数解问题作了大量的研究，但是许多求解方法往往是针对某个特定的方程，缺乏系统的理论和工具. 到18世纪末期，年轻的德国数学家Gauss在研究二平方和问题时，开创了一种有深远影响的新方法.

所谓**二平方和问题**即是：哪些自然数可以表示成两个整数的平方和？即：对哪些自然数 n，方程

$$x^2+y^2=n$$

有整数解 (x,y)？

例如在10以内：$1=1^2+0^2$，$2=1^2+1^2$，$4=2^2+0^2$，$5=2^2+1^2$，$8=2^2+2^2$，$9=3^2+0^2$，$10=3^2+1^2$，而3，6和7均不是二整数平方和.

Gauss的天才想法现在看来也许是很平凡的：他把这个纯粹是整数上的初等数论问题放到复数域中来考虑，在复数范围内，x^2+y^2 可以分解成

$$x^2+y^2=(x+\sqrt{-1}y)(x-\sqrt{-1}y)=|x\pm\sqrt{-1}y|^2.$$

于是，Gauss考虑形如 $x+\sqrt{-1}y$ 的复数，其中 $x,y\in Z$（即 x,y 为有理整数），这种数现在称为Gauss **整数**. 不难验证：两个Gauss整数相加、减、乘，仍为Gauss整数，用现在的语言，集合

$$Z[\sqrt{-1}]=\{x+\sqrt{-1}y\,|\,x,y\in Z\}$$

形成（交换）环，现在称为Gauss **整数环**. 如果我们定义

$$\begin{aligned}N(x+\sqrt{-1}y)&=(x+\sqrt{-1}y)(x-\sqrt{-1}y)\\&=|x\pm\sqrt{-1}y|^2=x^2+y^2\end{aligned}$$

为复数 $x+\sqrt{-1}y$ 的绝对值的平方，这叫作复数 $x+\sqrt{-1}y$ 的范数. 那末二平方和问题便可叙述成：

哪些自然数 n 是 Guass 整数的范数？

对于任意两个 Guass 整数 α 和 β，易知

$$N(\alpha\beta)=|\alpha\beta|^2=|\alpha|^2\cdot|\beta|^2=N(\alpha)N(\beta).$$

因此，若自然数 n 和 m 是 Gauss 整数 α 和 β 的范数，则 nm 是 Gauss 整数 $\alpha\beta$ 的范数. 这就表明：若 n 和 m 均是二整数平方和，则 nm 即是如此.

进一步，当 n 是素数 p 的情形，Gauss 证明了：素数 p 是二整数平方和的充分必要条件是 $p=2$ 或者 $p\equiv1(\mathrm{mod}4)$（即 p 是模 4 余 1 的奇素数）.（于是：素数 p 不为二平方和 $\Leftrightarrow p\equiv3(\mathrm{mod}4)$.）

为了对于任意的自然数 n 研究方程 $x^2+y^2=n$ 的整数解，Gausss 仔细考查了环 $\mathbf{Z}[\sqrt{-1}]$ 的算术特性. 他发现这个环也具有与有理整数环 $\mathbf{Z}$ 相类似的唯一分解特性. 即可以定义一些"Gauss 素数"，而每个 Gauss 整数均可以本质上唯一地表示成有限个 Gauss 素数之积，这样的交换环现在叫作唯一因子分解整环. 利用环 $\mathbf{Z}[\sqrt{-1}]$ 的唯一因子分解特性，Gauss 彻底解决了二平方和问题：

定理(Gauss) 自然数 n 可表成二整数平方和的充分必要条件是 $n=A^2B$，其中 A,B 是自然数，$B=1$ 或者 B 为有限个不同素数之积，并且 B 的每个素因子为 2 或模 4 余 1.

例 $n=728=2^3\cdot7\cdot13=2^2\cdot(2\cdot7\cdot13)$，由于 $7\equiv3(\mathrm{mod}4)$，可知 728 不是二整数平方和.

19 世纪中期，德国另一个数学家 Kummer 研究著名的 Fermat 猜想：当 $n\geqslant3$ 时，方程 $x^n+y^n=z^n$ 没有整数解 (x,y,z) 使得 $xyz\neq0$. 大家知道，为了证明这个猜想，我们只需对 $n=4$ 和 n 为奇素数 p 的情形证明即可. Fermat 本人用"无穷下降法"证明了 $n=4$ 的情形，而 $p=3$ 的情形由 Euler 和 Legendre 证明，$p=5$ 情形由

Dirichlet 证明. 其证明愈来愈复杂. Kummer 想到了采用 Gauss 的方法，即把这个纯粹初等数论问题放到复数域中考虑，令 $\zeta_p = e^{\frac{2\pi i}{p}}$，则 $x^p - z^p$ 在复数域中可以分解成 p 个数之积，而 Fermat 方程 $x^p + y^p = z^p$ 变成

$$(x - z)(x - \zeta_p z)(x - \zeta_p^2 z) \cdots (x - \zeta_p^{p-1} z) = - y^p.$$

注意到

$$1 + \zeta_p + \zeta_p^2 + \cdots + \zeta_p^{p-1} = \frac{1 - \zeta_p^p}{1 - \zeta_p} = 0,$$

不难验证，集合

$$\mathbf{Z}[\zeta_p] = \{a_0 + a_1\zeta_p + \cdots + a_{p-2}\zeta_p^{p-2} | a_0, \cdots, a_{p-2} \in \mathbf{Z}\}$$

形成环. Kummer 想当然地认为 $\mathbf{Z}[\zeta_p]$ 也是唯一因子分解整环，在这个假定下，他"证明"了 Fermat 猜想！但不久，他开始怀疑上述假定或许是不对的. 于是他作了大量的手工计算，证明了当 $3 \leqslant p \leqslant 19$ 时，$\mathbf{Z}[\zeta_p]$ 是唯一因子分解整环，从而他用统一的方法证明了当 $3 \leqslant p \leqslant 19$ 时，关于 $x^p + y^p = z^p$ 的 Fermat 猜想是对的. 另一方面，他发现环 $\mathbf{Z}[\zeta_{23}]$ 不是唯一因子分解整环，即其中存在数 α 具有本质上两种不同的素因子分解式. 此后，他竭力试图修补他的证明，于是创造了"理想数"概念，这便是近世代数中"环的理想"概念的起源.

交换环 R 中的子集合 I 叫作 R 的**理想**，是指满足以下两个条件：

(1) 若 $a, b \in I$，则 $a \pm b \in I$；

(2) 若 $a \in I, r \in R$，则 $ar \in I$.

例如：$\{0\}$ 和 R 本身均是 R 的理想，对每个元素 $\alpha \in R$，集合

$$(\alpha) \in \alpha R = \{\alpha r | r \in R\}$$

是 R 的理想. 这种理想叫**主理想**. 如果 R 中每个理想均是主理想，则 R 叫**主理想环**.

又如：设 A 和 B 均是环 R 的理想，则集合

$$\left\{\sum_i a_ib_i(\text{有限和})\,|\,a_i \in A, b_i \in B\right\}$$

也是R的理想，这个理想叫作A和B的乘积，表示成$AB(=BA)$.

环$Z[\sqrt{-1}]$不是域，即除法在$Z[\sqrt{-1}]$中不总是可进行的$\left(\frac{1+\sqrt{-1}}{2-\sqrt{-1}}=\frac{1}{5}-\frac{3}{5}\sqrt{-1}\notin Z[\sqrt{-1}]\right)$. 包含$Z[\sqrt{-1}]$的最小域是

$$Q(\sqrt{-1})=\{\alpha+\sqrt{-1}\beta\,|\,\alpha,\beta\in Q\}.$$

域$Q(\sqrt{-1})$是有理数域Q上的向量空间，维数是2，即$Q(\sqrt{-1})$是Q的二次扩域. 一般地，Q的每个二次扩域均叫作**二次（数）域**，二次域必可写成形式$Q(\sqrt{d})$，其中d是有理整数并且$\sqrt{d}\notin Z$. 类似地，包括环$Z[\zeta_p]$的最小域是

$$Q(\zeta_p)=\{\alpha_0+\alpha_1\zeta_p+\cdots+\alpha_{p-2}\zeta_p^{p-2}\,|\,\alpha_0,\cdots,\alpha_{p-2}\in Q\}.$$

可以证明这是Q的$p-1$次扩域，对于每个正整数$n\geqslant 3$，令$\zeta_n=e^{\frac{2\pi i}{n}}$，则包含$Q$和$\zeta_n$的最小域$Q(\zeta_n)$叫作**分圆域**，这是因为$1,\zeta_n,\zeta_n^2,\cdots,\zeta_n^{n-1}$恰好为复平面上单位圆周$|z|=1$的$n$个等分点.

一般的，有理数域的任意有限次扩域K叫作**代数数域**. 而K中元素α如果是某个多项式

$$f(x)=x^n+a_1x^{n-1}+\cdots+a_{n-1}x+a_n, a_i\in Z$$

（即$f(x)$的最高项系数为1，而其余系数均是有理整数）的根，则α叫作K中的**代数整数**. 例如当K是有理数域Q时，Q中的代数整数恰好就是有理整数. Q中整数全体形成环Z. 一般地，任意代数数域K中的代数整数全体也形成环，叫作K的**代数整数环**，表示成O_K. 要证明O_K是环（即K中两个代数整数的和、差、积仍是代数整数）并不是一件容易的事情，Dedekind 正是为了证明这件事，创造了代数学中又一个影响深远的概念 —— 模.

$Z[\sqrt{-1}]$和$Z[\zeta_n]$分别是域$Q(\sqrt{-1})$和$Q(\zeta_n)$的代数整

数环. 虽然像 $\mathbf{Z}[\zeta_{23}]$ 所表明的那样,代数整数环 O_K 不一定有元素的唯一因子分解性质,但是 Dedekind 证明了 O_K 有理想的唯一因子分解性质,即 O_K 中每个非零理想 $A(\neq O_K)$ 均可唯一地写成有限个素理想之乘积,这样的整环后人称之为Dedekind 整环. 可以证明

R 是主理想整环$\Rightarrow R$ 是唯一因子分解整环

$\Rightarrow R$ 是 Dedekind 整环.

进而,对于任意代数数域 K 的整数环 O_K,则第一个箭头的反向也成立,即 O_K 是唯一因子分解整环$\Leftrightarrow O_K$ 是主理想整环. 我们已经说过,O_K 是 Dedekind 整环,即具有理想的唯一因子分解性质. 自然我们要问:怎样判别 O_K 是否为主理想整环(即是具有元素唯一分解性质的环)?

O_K 中的理想对于乘法运算不形成群,我们要推广成分式理想,它可定义成

$$\frac{1}{n}A = \left\{ \frac{1}{n}\alpha \middle| \alpha \in A \right\},$$

其中 A 为 O_K 的理想,而 $n \in \mathbf{Z}$. 例如对 K 中每个元素 α,$\alpha O_K = \{\alpha a \mid a \in O_K\}$ 必是分式理想,这叫主分式理想. K 的所有非零分式理想形成交换群 I_K,而其中的主分式理想形成一个子群 P_K. 商群

$$C_K = I_K / P_K$$

叫作代数数域 K 的(分式) 理想类群. 代数数论的一个重要结果是:C_K 是有限交换群. 它的阶 $h_K = |C_K|$ 叫作 K 的理想类数(或简称类数). 于是:

O_K 为唯一因子整环 $\Leftrightarrow O_K$ 为主理想整环

$\Leftrightarrow O_K$ 的每个理想均是主理想

$\Leftrightarrow O_K$ 的每个分式理想均是主分式理想

$\Leftrightarrow P_K = I_K \Leftrightarrow C_K = \{1\} \Leftrightarrow h_K = 1.$

于是正整数 h_K 的大小可以作为Dedekind 整环 O_K 与唯一因子整环

相差程度的一个衡量标准.

现在我们可以说:Gauss 解决二平方和问题($x^2+y^2=n$) 正是使用了二次域 $Q(\sqrt{-1})$ 的类数为 1 这件事. 事实上,Gauss 研究了更一般的方程 $ax^2+bxy+cy^2=n$ 的整数解问题,它与一般二次域 $Q(\sqrt{d})$ 的类群和类数有密切联系. 另一方面,我们用 h_p 表示分圆域 $Q(\zeta_p)$ 的类数,则 Kummer 的前述结果可以说成:若 $h_p=1$,则关于 $x^p+y^p=z^p$ 的 Fermat 猜想是成立的. Kummer 仔细研究了分圆域 $Q(\zeta_p)$ 的性质,进一步证明了:对于每个奇素数 p,如果 $p\nmid h_p$,则关于 $x^p+y^p=z^p$ 的 Fermat 猜想是正确的. 在 100 以内的素数中,除了 $p=37$,59 和 67 之外,均满足 $p\nmid h_p$. 从而 Kummer 用深刻而统一的方法证明了:当 $3\leqslant p\leqslant 100$,$p\neq 37$,59,67 时,关于指数为素数 p 的 Fermat 猜想均成立. 这个结果与前人相比,无疑是一项光辉的成就.

正是由于 Gauss 和 Kummer 上述工作,开创了数论的一个重要的分支 ——(古典) 代数数论. 简言之,古典代数数论是研究任意代数数域 K 和它的整数环 O_K 的算术性质和代数性质的一门学问. 这一分支有许多重要的研究课题(如 O_K 的整基,单位群结构等),但是中心问题是研究代数数域 K 的理想类群 C_K(这是有限交换群) 的结构和类数 $h_K=|C_K|$ 的大小. 而二次域和分圆域是研究得最深刻的两类代数数域. 近 100 年来,代数数域得到丰硕的成果,但是仍有许多问题和猜想至今仍未解决. 关于二次域的类数问题,最著名的是 Gauss 的两个猜想:

(I) 只有有限多个虚二次域 $Q(\sqrt{d})$(即 $d<0$) 其类数为 1.

(II) 存在无限多个实二次域 $Q(\sqrt{d})$,(即 $d>0$) 其类数为 1.

Gauss 对虚二次域的类数作了大量的数值计算,他猜想类数为 1 的虚二次域只有 9 个,即 $d=-1,-2,-3,-7,-11,-19,-43,-67$ 和 -163 这 9 种情形. 直到 1967 年,美国数学

家 Stark 利用模形式理论才最终证明了这个猜想. 1971 年,Stark 和英国数学家,Fields 奖获得者 Baker 采用超越数论结果决定了类数为 2 的所有虚二次域. 1983 年,美国数学家 B. Gross 和 D. Zagier 对于模形式理论和椭圆曲线作了出色的工作,再结合 Goldfeld 在 1967 年的工作,得到了关于虚二次域类数问题的更一般结果,他们的工作获得 Cole 奖(这是数论和代数方面的最高奖). 另一方面,关于实二次域的 Gauss 猜想(Ⅱ),即是否存在无限多类数为 1 的实二次域,这问题至今未能解决. 1986 年 H. Cohn 和 Lenstra 提出一个更雄心勃勃的猜想,特别地,他们猜想:类数 1 的实二次域在所有实二次域中所占的比例约为 1/3. 这与目前算出的数据资料惊人地符合,但是离证明此猜想则相距更远了.

关于分圆域,Kummer 计算了一些分圆域 $\mathbb{Q}(\zeta_p)$ 的类数 h_p. 结果为 $h_3 = h_7 = h_{11} = h_{13} = h_{17} = h_{19} = 1$,而 $h_{23} = 3$. 他猜想:当素数 $p \geqslant 23$ 时,$h_p > 1$. 这个猜想于 1971 年才由日本数学家 Uchida 所证明. 1976 年,Masley 和 Montgomery 决定出类数为 1 的所有分圆域 $\mathbb{Q}(S_n)$(共 29 个). 这些结果均采用 §3 中所述的解析方法. 素数 p 叫作**正规**的,是指 $p \mapsto h_p$. 否则,p 叫作**非正规**的. 我们曾经提到 Kummer 的一个深刻的结果:对于正规素数 p,关于方程 $x^p + y^p = z^p$ 的 Fermat 猜想成立. 自然希望正规素数愈多愈好. 从概率论的角度考虑,人们猜想正规素数和非正规素数分别占素数的 $e^{-1/2} \approx 60.65\%$ 和 39.35%. 已经证明了不正规素数有无限多个. 但是人们目前还不清楚正规素数是否也有无限多个. 我们以 h_p^+ 表示分圆域 $\mathbb{Q}(\zeta_p)$ 的极大实子域 $\mathbb{Q}(\zeta_p + \zeta_p^{-1})$ 的理想类数. Kummer 证明了 $h_p^+ \mid h_p$,即 $h_p^- = h_p / h_p^+$ 是整数. Vandiver 猜想对每个素数 p,$p \mapsto h_p^+$. 关于 h_p^-,Kummer 猜想当素数 $p \to \infty$ 时,$h_p^- \sim 2p\left(\frac{p}{4\pi^2}\right)^{\frac{p-1}{4}}$. 这两个猜想至今仍未解决.

§ 3 Hilbert 的四个问题

> 我们当中有谁不想揭开未来的帷幕，看一看在今后的世纪里我们这门科学发展的前景和奥秘呢？我们下一代的主要数学思潮将追求什么样的特殊目标？在广阔而丰富的数学思想领域，新世纪将会带来什么样的新方法和新成果？
>
> Hilbert：《数学问题》，1900 年，巴黎国际数学家大会上的演说

自 Kummer 之后，Dedekind 对代数数域的一般性理论作了系统的代数整理. Dirichlet，Kronecker 为古典代数数论也作出重要贡献，而尤为值得一提的是 Hilbert 的工作. 1897 年，Hilbert 应德国数学会之邀写了著名的《数论报告》(Zahlenbericht). 这本书标志着古典代数数论的最终建立. 书中对于数域的二次扩张，分圆扩张和 Kummer 扩张的算术和代数性质都各写一章加以详细研究，为数域的一般 Abel 扩张积累了大量的材料和丰富的感性知识(数域扩张 L/K 叫作Abel 扩张，是指它是 Galois 扩张，并且其 Galois 群为 Abel 群). Hilbert 这项具有历史意义的工作导致他于 1898－1899 年提出如下的猜想：对于每个代数数域 K，以 H_K 表示 K 的极大不分歧 Abel 扩域，则 H_K/K 的 Galois 群同构于 K 的理想类群 C_K，并且 K 中每个理想在 H_K 中均为主理想. 第一个推断于 1907 年由 Hilbert 的学生 Furtwängler 证明. 现在把 H_K 叫作 K 的 Hilbert 类域. 后一推断(叫主理想定理)则是在日本数学家高木贞治于 1920 年建立了类域论之后，于 1930 年才由 Furtwängler 证明. 类域论是研究数域 K 的所有 Abel 扩张的漂亮的理论，它的中心议题是如何由 K 的自身特性来刻划 K 的所有 Abel 扩张. 例如，高木贞治证明了：K 的 Abel 扩域和 K 的“广义”类群是一一对应

的. 在这种对应下，K 的 Hilbert 类域恰好对应于 K 的通常的类群 C_K.

1900 年，Hilbert 在巴黎第二届国际数学家大会上应邀发表了著名的演说《数学问题》. 站在两个世纪交替的门坎上，Hilbert 首先讲了本节开头所引的那一段激动人心的话，随后他提出了著名的 23 个数学问题，这些问题的研究是本世纪数学发展的一个重要内容. 其中第 8，9，11，12 问题均是关于代数数论的. 现在谈谈 Hilbert 第 12 问题. 关于代数数论的另外三个 Hilbert 问题将在以后几节加以介绍.

1877 年，Kronecker 猜想：有理数域 $\mathbf{Q}$ 的每个（有限）Abel 扩张均是某个分圆域 $\mathbf{Q}(\zeta_m)$ 的子域. 这相当于说，所有分圆域的合成 $\bigcup\limits_{m\geqslant 2}\mathbf{Q}(\zeta_m)$ 是 $\mathbf{Q}$ 的最大 Abel 扩域 $\mathbf{Q}^{ab}$，其中 $\zeta_m = e^{\frac{2\pi i}{m}}$ 是指数函数 $f(z) = e^{2\pi iz}(z\in\mathbf{C})$ 在有理点 $z = 1/m$ 处的值. 因此，上述猜想又可叙述成：$\mathbf{Q}$ 的最大 Abel 扩域 $\mathbf{Q}^{ab}$ 是将指数函数 $e^{2\pi iz}$ 在所有有理点 $z\in\mathbf{Q}$ 处的值添加到 $\mathbf{Q}$ 上而得到的域. 这个猜想由 Weber 于 1886 年证明，现在称作 Weber-Kronecker 定理.

Kronecker 用类似的观点研究虚二次域 $K = \mathbf{Q}(\sqrt{d})(d<0)$ 的最大 Abel 扩域 K^{ab} 的构作问题. Kronecker 猜想（他把这叫作青春之梦，Jugendtraum）：虚二次域 K 的最大 Abel 扩域 K^{ab} 是将某些椭圆模函数在所有有理点上的值添加到 K 上而得到的域. 为了证明这个猜想，Weber 详细研究了椭圆模函数的性质. 但只是在建立了类域论之后，Kronecker 的青春之梦才变成了现实.

Hilbert 第 12 问题是：对任意代数数域 K，如何构作它的最大 Abel 扩域 K^{ab}？他猜想 K^{ab} 应当是将某些超越函数的特殊值添加到 K 上而得到的域. 这是一个非常困难的问题. 至今除了对 K 为有理数域和虚二次域得到圆满解决之外，其余结果甚少. 20 世纪中期，志村五郎（Shimura）对这个问题作了深刻的工作. 利用代数几

何与模形式理论，他对于一些实二次域 K，构作出最大 Abel 扩域 K^{ab}. 对于虚二次域 K，K^{ab} 的构作与具有复乘的椭圆曲线（一维交换代数群）的理论有密切联系. 而对于一类叫作复乘域（它是全实代数数域的全虚二次扩域）的 K，K^{ab} 的构作则与一般 Abel 簇（高维交换代数群）的算术理论有密切联系. 这使代数数论与代数几何及代数群的表示理论交织在一起，构成近代代数数论的一个十分重要而丰富的内容.

§4 Zeta 函数和 L- 函数 —— 解析理论

Zeta 函数知道代数数域的一切事情. 我们正是在劝说它把这些告诉给我们.

—— Harder（德国）

上一节谈到数论中采用代数工具，这一节则讲述数论中如何运用分析工具. 最早将分析工具用于数论的是 Euler. 他注意到，由自然数唯一表成有限个素数之积可知有恒等式

$$\sum_{n\geq 1} n^{-s} = \prod_{p} (1 - p^{-s}) \quad (s > 1).$$

但是当 $s \to 1$ 时，上式左边趋于 $+\infty$，由此证明存在无限多个素数 p. 德国大数学家 Riemann（1826 － 1866）把

$$\zeta(s) = \sum_{n\geq 1} n^{-s}$$

看成是复变量 $s = \sigma + it$ 的复值函数，并且详细研究了它的解析特性，和这些解析特性与素数分布的密切联系. 不难看出，Riemann zeta 函数 $\zeta(s)$ 是复半平面 $\sigma > 1$ 中的解析函数. Riemann 注意到 $\zeta(s)$ 在 $s = 1$ 处有 1 阶极点，并且留数 $\operatorname{res}_{s=1}\zeta(s) = \lim_{s\to 1}(s - 1)\zeta(s) = 1$. 利用这一事实他提出关于素数分布的一个猜想. 进而他证明了 $\zeta(s)$ 满足如下的函数方程

$$\xi(s) = \xi(1-s),$$

其中 $\xi(s) = \pi^{-s/2}\Gamma(s/2)\zeta(s)$，而 $\Gamma(s)$ 为 gamma 函数. 借助于这个函数方程，Riemann 把 $\zeta(s)$ 解析开拓成整个复平面上的亚纯函数，并且除了 $s=1$ 之外，$\zeta(s)$ 是解析的. 他还证明了 $s=-2,-4,-6,\cdots$ 是 $\zeta(s)$ 的零点. 进而，他猜想 $\zeta(s)$ 的其余零点（叫**非平凡零点**）均在直线 $\sigma = \frac{1}{2}$ 之上. 这个著名的 Riemann 猜想有许多重要的数论推论，但是 Riemann 猜想本身至今未能解决. Riemann 这篇发表于 1860 年的文章是他唯一的一篇关于数论的著作，但具有划时代的意义，因为它开创了用解析方法研究数论的一门学科——解析数论.

Dirichlet 考虑更一般的级数

$$L(s,\chi) = \sum_{n\geqslant 1}\chi(n)n^{-s} \quad (s=\sigma+it,\sigma>1),$$

其中 χ 是模 m 的**特征**. 这叫 Dirichlet$L$- 函数. 这个函数也有类似于 $\zeta(s)$ 的函数方程，并由此把 $L(s,\chi)$ 解析延拓到整个复平面上. Dirichlet 研究了这些 L- 函数的解析特性，并由此证明了：设 k 和 m 是互素的自然数，则算术级数 $k+ml(l=1,2,\cdots)$ 中存在无穷多个素数. 所谓**广义** Riemann **猜想**是：对每个模 m 的特征 χ，$L(s,\chi)$ 的非平凡零点均在直线 $\sigma=\frac{1}{2}$ 上.

现在设 K 是代数数域. 对于它的整数环 O_K 中的每个理想 A，商环 O_K/A 是有限的. 它的阶 $N(A) = |O_K/A|$ 叫作理想 A 的**范**. Dedekind 定义

$$\zeta_K(s) = \sum_{A} N(A)^{-s} \quad (s=\sigma+it,\sigma>1),$$

其中 A 过 O_K 的全体非零理想. 这叫作 K 的Dedekind zeta **函数**. 当 $\sigma>1$ 时，上式右边级数收敛，从而 $\zeta_K(s)$ 是右半平面 $\sigma>1$ 中的解析函数. 由于理想范是完全积性函数，即 $N(AB) = N(A)N(B)$，并且 O_K 是 Dedekind 整环（即每个理想唯一写成素理想之积），从

而当 $\sigma > 1$ 时有 Euler 乘积展开式

$$\zeta_K(s) = \prod_P (1 - N(P)^{-s})^{-1} \quad (\sigma > 1),$$

其中 P 过 O_K 的素理想. Dedekind 发现 $\zeta_K(s)$ 在 $s = 1$ 处有一阶极点，并且计算出留数值为

$$\operatorname*{Res}_{s=1}\zeta_K(s) = \lim_{s\to 1}(s - 1)\zeta_K(s) = \frac{2^{r_1}(2\pi)^{r_2}R_K h_K}{W_K |D_K|^{\frac{1}{2}}},$$

其中 $r_1 = r_1(K)$，$r_2 = r_2(K)$，R_K，W_K，D_K 和 h_K 均是与 K 有关的重要数值，反映 K 的代数特性，h_K 为 K 的理想类数. 于是我们看到，单单是 $\zeta_K(s)$ 的一个留数值便包含了代数数域 K 如此多的信息. 如果能够算出 $\operatorname*{Res}_{s=1}\zeta_K(s)$ 的值，则上式便给出计算理想类数 h_K 的解析公式. 对于分圆域 $K_m = Q(\zeta_m)$，借助于素数 p 在 K_m 中的分解规律，可知

$$\zeta_{K_m}(s) = \prod_\chi L(s, \chi^*),$$

其中 χ 过模 m 的所有特征，χ^* 是对应于 χ 的本原特征. 换句话说，分圆域的 Dedekind zeta 函数是已经熟知的 Dirichlet$L$- 函数之积，由此得到分圆域类数的解析公式. 事实上，用这种方法可得到任意 Abel 数域的类数解析公式. 德国数学家 Hasse 于 1952 年所写《Abel 域的类数公式》一书中，对于类数解析公式作了十分精细的研究.

Hibert 第 8 问题是素数分布问题和与此有关的 Riemann 猜想. 与此同时，Hilbert 还希望把这些结果用于任意代数数域中素理想分布问题上，以及研究 Dedekind zeta 函数 $\zeta_K(s)$ 的解析特性. 20 世纪 20 年代，Landau 和 Hecke 将研究素数分布的解析方法平行地移植到任意代数数域上，证明了素理想分布的类似结果. Hecke 还得到 $\zeta_K(s)$ 的函数方程，从而将 $\zeta_K(s)$ 解析延拓成整个复平面上的亚纯函数. 除了在 $s = 1$ 处有 1 阶极点之外，$\zeta_K(s)$ 在其他点均解析. 由函数方程不难决定，$\zeta_K(s)$ 在左半平面上的零点均在

某些负整数处，并且也不难决定这些零点的阶. 另一方面，$\zeta_K(s)$ 在右半平面中的零点都在带状区域 $0 \leqslant \sigma \leqslant 1$ 之内 $(s = \sigma + it)$. 猜想这些非平凡的零点均在直线 $\sigma = \frac{1}{2}$ 之上，这个广义 Riemann 猜想对于任何代数数域 K 都未能证明. 关于 Dedekind zeta 函数的另一个著名猜想是由 E. Artin 提出的：对于任意代数数域 E 和 F，如果 $E \subseteq F$，则 $\zeta_F(s)/\zeta_E(s)$ 是整个复平面上的解析函数. 1947 年，R. Brauer 利用群表示理论对于 F/E 是 Galois 扩张的情形证明了 Artin 猜想. 对于一般情形，这个问题还未解决. 设 $n \geqslant 2$，令

$$\zeta_F(s) = a_0(s-(1-n))^{r_n} + a_1(s-(1-n))^{r_n+1} + \cdots$$

是 $\zeta_F(s)$ 在 $s = 1-n$ 处的 Taylor 展开，其中 r_n 是 $\zeta_F(s)$ 在 $s = 1-n$ 处零点的阶. 20 世纪 70 年代，Lichtenbaum 猜想 a_0 和 K- 群 $K_n(O_F)$ 的阶之间有一个等式关系. 代数数域和代数 K- 理论之间的这一猜想的联系具有深刻的代数几何背景.

为了研究数论的各种问题，人们还构作了各种类型的 zeta 函数，例如：关于数域 Abel 扩张 E/F 的 Abel$L$- 函数 $L(s,E/F,\chi)$，关于数域 Galois 扩张 E/F 的 Artin$L$- 函数 $L(s,E/F,\chi)$（其中 χ 是 Galois 群 $\mathrm{Gal}(E/F)$ 某个表示的特征），关于有限域上代数簇 V 的 zeta 函数 $Z_V(s)$，等等. 研究这些 zeta 函数的解析特性（如解析延拓，函数方程等）在数论中具有重要的意义. 这一切表明，解析方法已成为现代数论一个重要的工具和内容.

§5 局部化和局部 —— 整体原则

除了代数和解析方法之外，代数数论的研究中还采用拓朴工具. 大家知道，有理数域 $\mathbb{Q}$ 中的绝对值 $|\ |$ 满足如下性质：对于有理数 a,b：

(1) $|a| \geqslant 0$；并且 $|a| = 0 \Leftrightarrow a = 0$；

(2) $|ab| = |a| \cdot |b|$;

(3)(三角形不等式) $|a+b| \leqslant |a| + |b|$.

这个绝对值使 Q 成为 Hausdorff 拓扑空间，两个有理数 a 和 b 的距离用数值 $|a-b|$ 来衡量. 这个拓扑空间不是完备的，它的完备化是实数域 R. 而每个实数均可表成十进小数形式

$$a_{-n}a_{-n+1}\cdots a_{-1}a_0.a_1a_2a_3\cdots = \sum_{l=-n} a_l \cdot 10^{-l},$$

其中 $0 \leqslant a_l \leqslant 9, a_l \in Z$.

19 世纪末，Hensel 受到解析函数在一点的 Taylor 展开式的启发，发现有理数域 Q 中不仅仅通常的绝对值满足上述三个条件. 取任一固定的素数 p 和任一实数 $\gamma, 0 < \gamma < 1$. 每个非零有理数 α 可唯一表示成

$$\alpha = p^{\lambda} \cdot \frac{a}{b},$$

其中 a 和 b 均是与 p 互素的整数，$b > 0$，并且 a 和 b 互素，而 λ 是整数. 我们令

$$|\alpha|_p = \gamma^{\lambda}.$$

而令 $|0|_p = 0$. 则不难看出 $|\ |_p$ 也满足上述三个条件. 事实上，它甚至于满足比(3)更强的条件

$$(3')\ |a+b|_p \leqslant \max(|a|_p, |b|_p).$$

有理数域对于 $|\ |_p$ 也是 Hausdorff 拓扑空间. $|\alpha|_p$ 叫作 α 的 p 进赋值. 对于 p 进赋值来说，有理数 α 中可被 p 的愈高幂次除尽，则 α 愈小. 例如对于 $p = 3$，则 $|3|_3 = \gamma > \gamma^2 = |9|_3$，$\left|-\frac{2}{3}\right|_3 = \gamma^{-1} > 1 = \gamma^0 = |2|_3$. 所以 p 进拓扑是与通常绝对值很不相同的拓扑. 对于不同的素数 p 和 q，p 进拓扑和 q 进拓扑也是不同的. 有理数域 Q 对于 p 进拓扑是不完备的，Q 对于 $|\ |_p$ 的完备化是 Q 的扩域，叫作 p 进数域，表示成 Q_p. Q_p 中每个元素叫作 p 进数，每个非零 p 进数 α 唯一表示成

$$\alpha = a_{-n} \cdot p^{-n} + a_{-n+1}p^{-n+1} + \cdots = \sum_{\lambda \geqslant -n} a_\lambda p^\lambda,$$

其中 $a_\lambda \in \{0, 1, \cdots, p-1\}$, $a_{-n} \neq 0$, $|\alpha|_p = \gamma^n$. 可以表成 $\alpha = \sum_{\lambda \geqslant 0} a_\lambda p^\lambda$(即 $|\alpha|_p \leqslant 1$)的 α 叫 p 进整数,它们形成环,叫 p 进整数环,表示成 $\mathbf{Z}_p$. 显然 $\mathbf{Z} \subset \mathbf{Z}_p$. $\alpha = \sum_{\lambda \geqslant 0} a_\lambda p^\lambda$ 是 $\mathbf{Z}_p$ 中可逆元素的充要条件是 $a_0 \neq 0$(即 $|\alpha|_p = 1$),这叫 p 进单位. 全体 p 进单位形成乘法群 U_p,叫 p 进单位群. 于是,每个非零 p 进数 $\alpha \in \mathbf{Q}_p$ 均唯一表成 $\alpha = p^n \cdot \varepsilon$,其中 $n \in \mathbf{Z}$, $\varepsilon \in U_p$. 今后记 $n = V_p(\alpha)$(于是 $|\alpha|_p = \gamma^{V_p(\alpha)}$,而 $\alpha \in \mathbf{Z}_p \Leftrightarrow V_p(\alpha) \geqslant 0$; $\alpha \in U_p \Leftrightarrow V_p(\alpha) = 0$).

可以证明,$\mathbf{Q}$ 中满足条件(1),(2),(3)的赋值本质上只有上述一些,即 p 进赋值 $|\ |_p$(p 过所有素数)和通常的绝对值 $|\ |$. 通常把 $|\ |$ 记成 $|\ |_\infty$,叫作无限赋值. 每个 $\mathbf{Q}_p$ 叫作 $\mathbf{Q}$ 在 p 处的局部域(包括 $\mathbf{Q}_\infty = \mathbf{R}$,这是 $\mathbf{Q}$ 在 $p = \infty$ 处的局部域).

设 K 是任意代数数域,P 是整数环 O_K 的一个素理想. 由于 O_K 是 Dedekind 整环,对于 K 中每个非零元素 α,主理想 $(\alpha) = \alpha O_K$ 可写成 $(\alpha) = P^n A$,其中 $n \in \mathbf{Z}$, A 是分子分母均与 P 互素的分式理想. 记 $V_P(\alpha) = n$, $|\alpha|_p = \gamma^n = \gamma^{V_P(\alpha)}$,而令 $|0|_P = 0$,则 $|\ |_P$ 是 K 的一个赋值(即满足条件(1),(2),(3). 事实上,$|\ |_p$ 满足(3′). 对不同的素理想 P, $|\ |_p$ 给出 K 的不同拓扑. 另一方面,K 还有一些无限赋值. 若 K 是 $\mathbf{Q}$ 的 n 次扩域,则 $K = \mathbf{Q}(\theta)$,其中 θ 是某个 n 次不可约多项式 $f(x) \in \mathbf{Q}[x]$ 的根. 则 K 中元素唯一表示成

$$C_0 + C_1\theta + \cdots + C_{n-1}\theta^{n-1} \quad (Ci \in \mathbf{Q}).$$

设 $f(x)$ 的 n 个根为

$$\theta_1, \cdots, \theta_{r_1}, \theta_{r_1+1}, \cdots, \theta_{r_1+r_2}, \theta_{r_1+r_2+1}, \cdots, \theta_{r_1+2r_2} \quad (r_1 + 2r_2 = n),$$

其中前 r_1 个为实根,后 $2r_2$ 个是 r_2 对共轭复根($\theta_{r_1+l} = \overline{\theta}_{r_1+r_2+l}$ ($1 \leqslant l \leqslant r_2$)),则对每个 i ($1 \leqslant i \leqslant n$),映射

$$\sigma_i : K \to C,$$

$$\sigma_i(C_0 + C_1\theta + \cdots + C_{n-1}\theta^{n-1}) = C_0 + C_1\theta_i + \cdots + C_{n-1}\theta_i^{n-1}$$

是 K 到 C 中的单同态(嵌入). 对于 K 中元素 α,令

$|\alpha|_{\infty_i} = |\sigma_i(\alpha)|$(其中 $|\ |$ 是复数通常绝对值),则 $|\ |_{\infty_i}(1 \leqslant i \leqslant r_1 + r_2)$ 均是 K 的赋值,它们叫 K 的无限赋值. 可以证明:代数数域 K 本质上只有 P 进赋值 $|\ |_P$(P 过 O_K 的所有素理想) 和上述 $r_1 + r_2$ 个无限赋值. K 对于赋值 $|\ |_P$ 的完备化叫作 K 在 P 处的局部域,表示成 K_P. 注意 $K_{\infty_i} = R$, $(1 \leqslant i \leqslant r_1)$, $K_{\infty_i} = C$ $(r_2 + 1 \leqslant i \leqslant r_1 + r_2)$.

一般来说,由于局部域 K_P 是完备的拓朴域,K_P 中的数论问题(局部问题) 比 K 中(整体问题) 要容易处理. 通过局部域 K_P 来解决 K 中的数论问题是代数数论的一个重要方法,叫局部化方法. 例如:设 L/K 是数域的扩张,$L = K(\theta)$, $f(x)$ 是 θ 在 K 上的极小多项式,则 O_K 中每个素理想 P 在 O_L 中的分解情形,由 $f(x)$ 在局部域 K_P 中的分解所完全决定,即这是只与一个 K_P 有关的性质,这样的性质叫局部性质. 还有一些关于 K 的数论问题,由它在所有局部域 K_P 上的解决所决定,这种解决方法叫作局部-整体原则(即通过所有局部来把握整体),或叫 Hasse 原则,这是由于 Hasse 在研究 Hilbert 第 11 问题时,给出这种原则的第一个精彩的例子.

Hilbert 第 11 问题是求系数属于某代数数域 K 的多变量二次方程在 K 中的解. 通过配方可以消去方程的一次项,从而方程可以化为 $f(x_1, \cdots, x_n) = \alpha (\alpha \in K)$, 其中 $f(x_1, \cdots, x_n) = \sum_{i,j=1}^{n} a_{ij}x_ix_j$ 是系数 a_{ij} 属于 K 的二次型.

一般的,设 $g(x_1, \cdots, x_n)$ 是系数属于 K 的任意多项式. 由 $K \subset K_P$ 可知:若方程 $g(x_1, \cdots, x_n) = 0$ 在 K 中有解,则它在每个局部域 K_P 中均有解. 但其逆命题通常不成立. 1923 年,Hasse 对于二次型的情形证明了如下的漂亮结果:设 $f(x_1, \cdots, x_n) =$

$\sum_{i,j=1}^{n} a_{i,j}x_ix_j, a_{ij} \in K$,则

(1) 对每个 $0 \neq \alpha \in K$,方程 $f(x_1, \cdots, x_n) = \alpha$ 有整体解(即在 K 中有解)$\Leftrightarrow$ 该方程局部处处有解(即在每个局部域 K_P 中有解).

(2) $f(x_1, \cdots, x_n) = 0$ 有整体的非零解 $\Leftrightarrow f(x_1, \cdots, x_n) = 0$ 在每个局部域 K_P 中均有非零解.

Hasse 还得到关于二次型的另一些局部-整体结果. 现在我们以 $K = \mathbb{Q}$ 举例说明如何利用 Hasse 原则. 设 a,b 是 $\mathbb{Q}_p$ 中非零元素,定义

$$\left(\frac{a,b}{p}\right) = \begin{cases} 1, \text{如果 } z^2 - ax^2 - by^2 = 0 \text{ 在 } \mathbb{Q}_p \text{ 中有非零解} (z,a,y); \\ -1, \text{否则}. \end{cases}$$

这叫 Hilbert 符号,显然

$$\left(\frac{a,b}{p}\right) = \left(\frac{b,a}{p}\right), \left(\frac{ac^2,b}{p}\right) = \left(\frac{a,b}{p}\right) (\text{其中 } 0 \neq c \in \mathbb{Q}_p).$$

$$\left(\frac{a,b}{\infty}\right) = \begin{cases} -1 \text{ 若 } a < 0, b < 0; \\ 1, \text{否则}. \end{cases}$$

对于素数 p,设 $a = p^\alpha u, b = p^\beta v, u, v \in U_p$,则 Hilbert 符号可用下列公式来计算:对于奇数 p,

$$\left(\frac{a,b}{p}\right) = (-1)^{\alpha\beta\varepsilon(p)} \left(\frac{u}{p}\right)^\beta \left(\frac{v}{p}\right)^\alpha,$$

而

$$\left(\frac{a,b}{2}\right) = (-1)^{\varepsilon(u)\varepsilon(v) + \alpha w(v) + \beta w(u)},$$

其中 $\left(\frac{u}{p}\right)$ 为 Legendre 符号. $\varepsilon(u) = \frac{u-1}{2}, w(u) = \frac{u^2-1}{8}$. 对于对角二次型 $f(x_1, \cdots, x_n) = a_1x_1^2 + \cdots + a_nx_n^2, a_i \in \mathbb{Q} - \{0\}$. 定义

$$\varepsilon_P(f) = \prod_{1 \leqslant i < j \leqslant n} \left(\frac{a_i, a_j}{p}\right).$$

设 a_1,a_2,a_3,a 是 $\mathbf{Q}_p$ 中非零元素,则方程 $a_1x_1^2+a_2x_2^2+a_3x_3^2=a$ 在 $\mathbf{Q}_p$ 中有解的充分必要条件是:或者 $-aa_1a_2a_3$ 为 $\mathbf{Q}_p$ 中非平方元素,或者 $-aa_1a_2a_3$ 是 $\mathbf{Q}_p$ 中平方元素并且 $\left(\frac{-1,\ -a_1a_2a_3}{p}\right)=\varepsilon_p(f)$. 特别地:对于 $a\in\mathbf{Z}$, 则 $x_1^2+x_2^2+x_3^2=a$ 在 $\mathbf{Q}_p$ 中有解的充分必要条件是:或者 $-a$ 是 $\mathbf{Q}_p$ 中非平方元素,或者 $-a$ 是 $\mathbf{Q}_p$ 中平方元素并且 $\left(\frac{-1,\ -1}{p}\right)=1$. 注意当 p 为奇素数时, $\left(\frac{-1,\ -1}{p}\right)=1$. 从而 a 在 $\mathbf{Q}_p$ 中总可表成三个数的平方和. 由于 $\left(\frac{-1,\ -1}{\infty}\right)=-1$, 从而只有非负整数 a 才可表成三个实数的平方和. 最后, $\left(\frac{-1,\ -1}{2}\right)=-1$. 因此 a 在 $\mathbf{Q}_2$ 中可表成三平方和的充要条件是 $-a$ 为 $\mathbf{Q}_2$ 中非平方元素. 于是由 Hasse 原则:正整数 n 是三个有理数平方和的充要条件为 $-n$ 是 $\mathbf{Q}_2$ 中非平方元素,即 $n=4^a\cdot b$, 其中 $4\nmid b$, $b\not\equiv 7 \pmod 8$. 不难证明:若 n 可表为三个有理数平方和,则必可表为三个整数的平方和. 于是我们得到

Gauss 定理 正整数 n 可表为三整数平方和的充分必要条件是 $n=4^a\cdot b$, 其中 $b\not\equiv 7 \pmod 8$, $4\nmid b$. 由此不难证明:

Lagrange 定理 每个正整数均可表成四个整数的平方和.

一般说来,研究方程 $\sum_{i,j=1}^{n}a_{ij}x_ix_j=a\,(a_{ij},a\in\mathbf{Z})$ 的整数解和解数公式,则是比求域中解更为困难的数学问题. 研究这问题的重要工具是模形式理论. 另一方面,对于一般的高次多项式方程没有 Hasse 的局部整体原则,所以研究它在某代数数域 K(或其整数环 O_K)中的解是很困难的. 但是对于形如 $y^2=x^3+ax+b$ 的方程已经有很丰富的理论,形成数论一个重要分支:椭圆曲线理论. 下节我们扼要介绍模形式和椭圆曲线这两个有密切联系的数论分支.

§ 6　模形式和椭圆曲线

设 $f(x_1,\cdots,x_n)=\sum_{i,j=1}^{n}a_{ij}x_ix_j$ 是正定二次型，其中 $a_{ij}\in Z$. 对于每个整数 $m\geqslant 0$，$f(x_1,\cdots,x_n)=m$ 有有限多整解 $(x_1,\cdots,x_n)\in Z^n$. 整解个数叫作 f 表示 m 的表法数，表示成 $r(f,m)$. 它的母函数为

$$\theta_f(z)=\sum_{m=0}^{\infty}r(f,m)e^{m\pi iz}=\sum_{x_1,\cdots,x_n\in Z}e^{\pi izf(x_1,\cdots,x_n)}.$$

由于 f 是正定二次型，当 z 属于上半平面 $H=\{z=x+iy|y>0\}$ 时上面级数收敛，因此 $\theta_f(z)$ 是 H 中解析函数，叫作关于二次型 f 的 theta 函数. 于是，研究 $r(f,m)$ 的值归结于研究 $\theta_f(z)$ 的性状. 事实上，对于满足一定条件的二次型 f，$\theta_f(z)$ 是某种类型的模形式.

我们以 $SL_2(Z)$ 表示行列式为 1 的整系数 2 阶方阵全体形成的乘法群. 则 H 上解析函数 $f(z)$ 叫作权为 $2k$ 的模形式，是指它满足如下两条件：

(I)　$f\left(\frac{az+b}{cz+d}\right)=(cz+d)^{2k}f(z)$，对每个 $\begin{pmatrix}ab\\cd\end{pmatrix}\in SL_2(Z)$，

特别取 $\begin{pmatrix}ab\\cd\end{pmatrix}=\begin{pmatrix}11\\01\end{pmatrix}$，则 $f(z+1)=f(z)$，即 f 是周期为 1 的函数，从而有 Fourier 展开. 我们还要求展开式没有"负"项，即

(Ⅱ)　$f(z)=\sum_{m=0}^{\infty}a_mq^m\quad(q=e^{2\pi iz})$.

模形式的例子：

例 1(Eisenstein 级数)　设 k 为正整数，

$$G_k=G_k(z)=\sum_{\substack{m,n\in Z\\(m,n)\neq(0,0)}}(mz+n)^{-2k}.$$

当 $k\geqslant 2$ 时，此级数在 H 中收敛. 不难证明 $G_k(z)$ 是权 $2k$ 的模形

式，它的 Fourier 展开式为

$$G_k(z) = 2\zeta(2k) + 2 \cdot \frac{(2\pi i)^{2k}}{(2k-1)!} \sum_{m=1}^{\infty} \sigma_{2k-1}(m) q^m,$$

其中 $\zeta(s)$ 为 Riemann zeta 函数，$\sigma_k(m) = \sum_{d|m} d^k$（即 m 的全部正因子的 k 次幂之和）.

若 $F(z)$ 和 $E(z)$ 分别为权 k, k' 的模形式，则 $F(z)E(z)$ 是权 $k+k'$ 的模形式. 令 M_k 为所有权 k 的模形式全体，这是 $\mathbf{C}$ 上的向量空间. 而 $M = \sum_{k=0}^{\infty} M_k$ 是分次的代数. 可以证明 $M = \mathbf{C}[G_2, G_3]$. 于是对每个 $k \geqslant 0$，$\{G_2^\alpha G_3^\beta \mid 2\alpha + 3\beta = k, \alpha, \beta \in \mathbf{Z}, \alpha, \beta \geqslant 0\}$ 是向量空间 M_k 的一组基. 特别地

$$\dim M_k = \sum_{\substack{\alpha,\beta \geqslant 0 \\ 2\alpha + 3\beta = k}} 1 = \begin{cases} \left[\frac{k}{6}\right], \text{若 } k \equiv 1 \pmod{6}; \\ \left[\frac{k}{6}\right] + 1, \text{否则}. \end{cases}$$

例 2 （二次型的 theta 函数） 设 $f(x_1, \cdots, x_n) = \sum_{i,j=1}^{n} a_{ij} x_i x_j$，其中 (a_{ij}) 是正定对称整阵，并且主对角元素 a_{ii} $(1 \leqslant i \leqslant n)$ 均为偶数，而 $\det(a_{ij}) = 1$. 可以证明，$8|n$，并且 $\theta_f(z)$ 是权 $n/2$ 的模形式. 例如对 $n = 8$ 时，

$$(a_{ij}) = \begin{bmatrix} 2 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & -1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 \end{bmatrix}$$

满足上述条件. 从而 $\theta_f(z) \in M_2 = \mathbf{C}G_2$. 因此 $\theta_f(z) = \alpha G_2(z)$，其

中 $\alpha \in C$. 由于当 $2 \nmid m$ 时 $r(f,m) = 0$. 而 $r(f,0) = 1$. 因此

$$\theta_f(z) = 1 + \sum_{m \geqslant 1} r(f,2m) q^m,$$

$$G_2(z) = 2\zeta(4) + 2\frac{(2\pi i)^4}{3!}\sum_{m \geqslant 1}\sigma_3(m) q^m,$$

$$\zeta(4) = \frac{\pi^4}{90}.$$

可知 $\theta_f(z) = \frac{1}{2\zeta(4)} G_2(z)$, 从而 $r(f,2m) = 240\sigma_3(m)$ (对 $m \geqslant 1$).

为了适用于更多的二次型,我们还要研究更一般的模形式(例如条件(Ⅰ)只对 $SL_2(\mathbf{Z})$ 的某个子群成立,等等). 以 $r_n(m)$ 表示 m 表成 n 个整数平方和的表法数,令 $\theta(z) = \sum_{m \in \mathbf{Z}} q^{m^2} = 1 + 2\sum_{m=1}^{\infty} q^{m^2}$, 则 $\theta^n(z) = \sum_{m=0}^{\infty} r_n(m) q^m$. 当 n 是偶数并且 $n \geqslant 4$ 时, $\theta^n(z)$ 是某种类型的权 $n/2$ 的模形式,利用这种模形式空间的基,并将 $\theta^n(z)$ 表成基的线性组合,比较 Fourier 系数可得到

$$r_4(m) = 8\sum_{\substack{d \mid m \\ 8 \nmid d}} d, \quad r_8(m) = 16\sum_{d \mid m}(-1)^{m-d} d^3,$$

等等. 而对于 n 为奇数的情形,为了得到 $r_n(m)$ 的公式,需要研究半整权的模形式. 为了研究用一个二次型表示另一个二次型的问题,则有模形式的高维推广,叫作 Siegel 模形式. 另一种推广是将 $\mathbf{Q}$ 改成任意代数数域 K,即研究系数属于 O_K 的二次型表示 O_K 中数的问题,这种推广叫Hilbert 模形式. 上述各种类型的模形式全体均形成 $\mathbf{C}$ 上有限维向量空间.

模形式理论的最基本问题是:计算各种类型模形式空间的维数,给出它们的一组基,研究模形式 Fourier 系数的数论意义和性质(整性,渐近估计,同余性质等). 这个基本问题目前也只对少数情形得到满意的解决. 模形式理论是现代数论一个非常活跃的分支,它与许多其他分支都有紧密的联系.

给了两个复数 g_2 和 g_3，$g_2^3 - 27g_3^2 \neq 0$，则存在 $\omega \in H$，使得 $60G_2(\omega) = g_2$，$140G_3(\omega) = g_3$. 令

$$\wp(z) = \wp(z,\omega) = \sum_{\substack{m,n\in Z \\ (m,n)\neq(0,0)}} \left[(z-(m+n\omega))^{-2} - (m+n\omega)^{-2}\right]$$

$$= z^{-2} + \sum_{k=1}^{\infty}(2k+1)G_{k+1}(\omega)z^{2k},$$

则 $(x,y) = (\wp(z),\wp'(z))$ 满足方程

$$E: y^2 = 4x^3 - g_2x - g_3.$$

这个方程定义的曲线叫作**复椭圆曲线**. $\wp(z) = \wp(z,\omega)$ 叫作 Weierstrass $\wp$ 函数，它是双周期的复值亚纯函数，即 $\wp(z) = \wp(z+1) = \wp(z+\omega)$. 并且 $\wp$ 在 $z=0$ 处有 2 阶极点. 并且我们有一一对应

$$C/Z + wZ \to E, z \mapsto (\wp(z),\wp'(z)),$$

其中 0 在 E 中的象看成是曲线上的“无穷远点”. 由于 $C/Z+\omega Z$ 自然是加法交换群，从而通过上述对应关系可将曲线 E 的所有复解 (x,y) 之间引入加法运算，使之构成交换群，零元素为无穷远点. 于是，我们由模形式转而介绍现代数论的另一分支 —— 椭圆曲线算术理论.

设 K 为域，$a,b \in K$，$4a^3 + 27b^2 \neq 0$，则曲线

$$E: y^2 = x^3 + ax + b$$

叫作**定义于** K **上的椭圆曲线**. 我们以 $E(K)$ 表示曲线 E 的所有 K 中的解 $(x,y)(x,y\in K)$ 而构成的集合(包括无穷远点 0)，则可定义加法运算使 $E(K)$ 为交换群. 当 K 是代数数域时，著名的 Mordell-Weil 定理是说，群 $E(K)$ 是有限生成的，从而可分解为：

$$E(K) = E(K)_f + E(K)_t (\text{直和}),$$

其中 $E(K)_t$ 是 $E(K)$ 的有限阶元素的构成的子群，这是有限交换群. 而 $E(K)_f$ 同构于 Z^r，$r = \text{rank}(E(K))$ 叫作群 $E(K)$ 的**秩**.

椭圆曲线算术理论的最基本问题是：对于给定的代数数域 K

和定义于 K 上的椭圆曲线 E，决定 $E(K)t$（或它的结构）和计算 $\mathrm{rank}(E(K))$，并且给出 $E(K)_f$ 的一组基.

对于每个具体的曲线 E，决定 $E(K)_t$ 不太困难，这可通过研究 $E(K_p)$（对每个局部化 K_p）的结构而解决. 但是决定 $E(K)_f$ 是十分困难的. 关于 $E(K)$ 的结构有如下两个重要猜想，我们令

$$T(K)=\{E(K)_t \mid E \text{ 为定义于 } K \text{ 上的椭圆曲线}\}.$$

猜想 1 对于每个代数数域 K，$T(K)$ 是有限集合，换句话说，只有有限多个有限交换群可作为定义于 K 上某椭圆曲线的 torsion 部分.

这个猜想目前只对 $K=\mathbf{Q}$ 的情形被解决. 1978 年，哈佛大学的 B. Mazur 证明了：$T(\mathbf{Q})$ 由下列 15 个群组成：$\mathbf{Z}/(N\mathbf{Z})$ $(1\leqslant N\leqslant 10, N=12)$，$\mathbf{Z}/(2\mathbf{Z})\oplus\mathbf{Z}/(2N\mathbf{Z})$ $(1\leqslant N\leqslant 4)$. 1969 年，苏联数学家 Manin 证明了：对于每个代数数域 K 和每个素数 p，集合 $\{E(K)_t$ 的 p-sylow 子群 $\mid E$ 为定义于 K 上椭圆曲线$\}$ 是有限的，但是猜想 1 对于每个 $K\neq\mathbf{Q}$ 目前均未解决[后记 1].

猜想 2 $\{\mathrm{rank}(E(\mathbf{Q}))\mid E$ 为定义于 $\mathbf{Q}$ 的椭圆曲线$\}$ 是无限集合. 换句话说，定义于 $\mathbf{Q}$ 上的椭圆曲线的秩可以任意大. 1985 年，J. F. Mestre 构作出 $\mathrm{rank}(E(\mathbf{Q}))\geqslant 14$ 的 $\mathbf{Q}$ 上椭圆曲线 E，但是猜想 2 本身仍未解决.

设 E 是有限域 F_q 上的椭圆曲线，则对于每个 $n\geqslant 1$，E 也是 F_{q^n} 上的椭圆曲线，以 N_n 表示 $|E(F_{q^n})|$，即 E 在 F_{q^n} 中的解数（包括无穷远点）. 母函数

$$Z_E(T)=\exp\Big(\sum_{n=1}^{\infty}N_nT^n/n\Big)$$

叫作**有限域 F_q 上椭圆曲线 E 的 zeta 函数**. 1935 年，Hasse 和 Davenport 证明了：

(1) $Z_E(T)=\dfrac{1-aT+qT^2}{(1-T)1-qT)}$，其中 $a=(1+q)-N_1$；

(2)(函数方程)$Z_E\left(\frac{1}{qT}\right) = Z_E(T)$;

(3)$1 - aT + qT^2 = (1 - \alpha T)(1 - \bar{\alpha}T)$,其中 $|\alpha| = \sqrt{q}$;

(4) 对每个 $n \geqslant 1, N_n = (1 + q^n) - (\alpha^n + \bar{\alpha}^n)$.(于是由(3)可知 $|N_n - (1 + q^n)| \leqslant 2q^{n/2}$.)

如果令 $T = q^{-s}$,并记 $\zeta_E(s) = Z_E(q^{-s})$,则(2)中函数方程可写成 $\zeta_E(1 - s) = \zeta_E(s)$. 而(3)可以叙述成:$\zeta_E(s)$ 的两个零点均在直线 $R_e(s) = \frac{1}{2}$ 上,这正是关于 Riemann zeta 函数 $\zeta(s)$ 的 Riemann 猜想在有限域上的模拟,我们在下节还要回到这个议题中来. 最后,(4)式的意义是:由 E 在 $\mathbf{F}_q$ 中的解数 N_1 可完全决定 E 在所有扩域 $\mathbf{F}_{q^n}(n \geqslant 1)$ 中的解数 N_n.

现在设 $E: y^2 = x^3 + ax + b$ 是代数数域 K 上的椭圆曲线,于是 $4a^3 + 27b^2 \neq 0$,从而对 O_k 中几乎所有(即除有限个之外)的素理想 P,a 和 b 均是 P 进整数,并且 $4a^3 + 27b^2 \notin P$. 对于这种 P,通过将系数模 P 之后,E 变成有限域 $\mathbf{F}_q = O_K/P$ 上的椭圆曲线 $E_p(q = q(P))$,叫作 E 在 P 处有**好的约化**. 我们以 $L_P(s)$ 表示有限域 $\mathbf{F}_q$ 上 E_P 的 zeta 函数 $\zeta_{E_p}(s)$ 的分子 $1 - a_P q(P)^{-s} + q(P)^{1-2s}$. 对于使 E 有坏约化的 P,我们也适当定义 $L_P(s)$(详情从略). 将所有这些局部因子收集在一起,得到的整体函数

$$L_E(s) = \prod_P L_p(s)^{-1} = \sum_{n=1}^{\infty} c_n n^{-s}$$

叫作**K 上椭圆曲线 E 的 L-函数**. 由前面(4)中对 N_n 的估计可知这个级数在半平面 $R_e(s) > 3/2$ 中收敛. 猜想 $L_E(s)$ 可解析开拓成整个复平面上并且满足某种函数方程. 我们以 $K = \mathbf{Q}$ 为例,设 E 是 $\mathbf{Q}$ 上的椭圆曲线,$L_E(s) = \sum_{n=0}^{\infty} c_n n^{-s}$,则有如下的

谷山-Weil 猜想:

(1)(函数方程)令 $\zeta_E(s) = N^{s/2}(2\pi)^{-s}\Gamma(s)L_E(s)$,其中 N 是曲

线 E 的"导子"(Conductor),则 $\zeta_E(s) = \pm \zeta_E(2-s)$;

(2) 级数 $L_E(s) = \sum c_n n^{-s}$ 所对应的函数 $f(z) = \sum_{n\geqslant 1} c_n e^{2\pi i n z}$ 是关于同余子群 $\Gamma_0(N)$ 的权 2 模形式.

对于许多种情形均已验证这个猜想是对的. 但是并未完全解决. 1986 年, Frey 证明了:若谷山-Weil 猜想成立,则 Fermat 猜想便成立. 这个结果再次使人们对数论各分支之间的紧密联系惊叹不已[后记 2].

正像代数数域 K 的 zeta 函数 $\zeta_K(s)$ 的解析特性反映出域 K 的算术性质一样,人们猜想 $L_E(s)$ 的解析特性也反映出群 $E(\mathbf{Q})$ 的结构特性. 最吸引人的猜想是:

(Birch, Swinnerton-Dyer) BSD 猜想: 设 E 是 $\mathbf{Q}$ 上的椭圆曲线,则 $L_E(s)$ 在 $s=1$ 处零点的阶 $\operatorname{ord}_{s=1} L_E(s)$ 等于 $r = \operatorname{rank}(E(\mathbf{Q}))$, 并且 $L_E(s)$ 在 $s=1$ 处 Taylor 展开式的第一个系数 $\lim_{s\to 1}(s-1)^{-r} L_E(s)$ 也可用 E 的一些数论特性来表达(详情从略).

1977 年, Coates 和 Wiles 证明了: $L_E(1) \neq 0$, (即 $\operatorname{ord}_{s=1} L_E(s) = 0) \Rightarrow \operatorname{rank} E(\mathbf{Q}) = 0$ (即 $E(\mathbf{Q})$ 是有限群). 1986 年, B. Gross 和 D. Zagier 证明了: $\operatorname{ord}_{s=1} L_E(s) = 1 \Rightarrow \operatorname{rank}(E(\mathbf{Q})) \geqslant 1$. 1987 年, K. Rubin 对于有复乘的椭圆曲线 E, 证明了: 当 $\operatorname{ord}_{s=1} L_E(s) \leqslant 1$ 时, $\operatorname{rank}(E(\mathbf{Q})) = \operatorname{ord}_{s=1} L_E(s)$. 椭圆曲线理论的这些最近成果使用了分圆域理论、模形式理论、形式群、P 进分析和代数几何等方面的许多成就. 但目前 BSD 猜想离完全解决似乎还相距很远[后记 3].

椭圆曲线是一维的交换代数群. 而高维交换代数群即 Abel 簇的算术理论目前也是一个活跃的数论分支.

§7 代数几何的算术理论

代数几何的基本任务是研究多项式方程(组) 在某个域或环

上的解与解集合的特性，研究的内容和手段有几何、分析和数论（即算术）三种形式. 这里介绍代数几何的算术理论.

首先应当介绍最近一项重大成果. 设 $C: f(x,y)=0$ 是代数数域 K 上一条代数曲线（即 $f(x,y)\in K[x,y]$），$g=g(C)$ 是曲线 C 的亏格（genus，这是曲线的双有理不变量）. 当 $g=0$ 时，f 是次数 $\leqslant 2$ 的多项式，研究 C 的解是容易的. 当 $g=1$ 时，C 即是椭圆曲线. 这时 $C(K)$ 是有限生成交换群. 特别的，C 在 K 中有无限多解的充要条件是 $\mathrm{rank}(C(K))\geqslant 1$（猜想这又等价于 $L_C(1)=0$）. 1922 年，英国数学家 Mordell 提出一个大胆的猜想：当 $g(C)\geqslant 2$ 时，$C(K)$ 有限，即曲线 C 在代数数域上只有有限多点. 人们对这个猜想作了 60 年的努力，直到 1983 年，当时 28 岁的德国人 Faltings 利用代数几何与数论的最新成果一举证明了这个猜想. 由于这项工作，他获得 1986 年的 Fields 奖.

为了说明这项结果的威力，我们只需提一下它的一个小小的推论. 熟知 Fermat 曲线 $X^n+Y^n=1$ 的亏格是 $g=\frac{1}{2}(n-1)(n-2)$. 特别当 $n\geqslant 4$ 时，$g\geqslant 3$. 因此由 Faltings 定理可知当 $n\geqslant 4$ 时，$X^n+Y^n=1$ 只有有限多有理数解. 而 Fermat 猜想是说对每个 $n\geqslant 3$，$x^n+y^n=z^n$ 没有整数解 (x,y,z) 使得 $xyz\neq 0$. 注意若 (x,y,z) 是 $x^n+y^n=z^n$ 的解，则对每个 $0\neq a\in Z$，(ax,ay,az) 也是它的解，这样两个解称作是本质相同的. 不难看出，$x^n+y^n=z^n$ 的本质不同的解 $(x,y,z)(xyz\neq 0)$ 和 $X^n+Y^n=1$ 的有理数解 (X,Y) 是一一对应的 $\left(取 X=\frac{x}{z}, Y=\frac{y}{z}\right)$. 于是由 Faltings 结果推出：对每个 $n\geqslant 4$，$x^n+y^n=z^n$ 只有有限多本质不同的整数解. 这是 Fermat 猜想的一个重大进展.

有限域上的代数几何也具有丰富内容、优美理论和广泛前景. 我们这里只介绍 Weil 猜想. 我们在上节介绍过有限域上椭圆曲线的 Hasse-Davenport 结果. 1941 年法国数学家 A. Weil 在第二次世

界大战期间移居美国时，对于有限域上一般的代数曲线提出下列猜想.

设 $f(x,y) \in \mathbf{F}_q[x,y]$，$C: f(x,y) = 0$ 是绝对不可约的非奇异代数曲线，$g = g(C)$ 为 C 的亏格. N_n 是该曲线在有限域 $\mathbf{F}_{q^n}$ 中的（射影）解数. $Z_C(T) = \exp\left(\sum_{n\geq 1} N_n T^n/n\right)$ 为曲线 C 的 zeta 函数. 则

(1)（有理性）$Z_C(T) = \dfrac{F(T)}{(1-T)(1-qT)}$，其中 $F(T) \in \mathbf{Z}[T]$；

(2)（函数方程）$Z_C\left(\dfrac{1}{gT}\right) = q^{g-1}T^{2g-2}Z_C(T)$；

(3)（Riemann 猜想的模拟）$F(0) = 1$，$\deg F = 2g$，并且 $F(T) = \prod_{i-1}^{2g}(1 - w_i T)$，其中 $|w_i| = \sqrt{q}\ (1 \leqslant i \leqslant 2g)$；

(4)（N_n 的估计）　$N_n = 1 + q^n - \sum_{i=1}^{2g} w_i^n$. 特别地，$|N_n - (1 + q^n)| \leqslant 2gq^{n/2}$.

1924 年，E. Artin 对于曲线 $y^2 = f(x)$，$f(x) \in \mathbf{F}_q[x]$，$2 \nmid q$ 证明了上述猜想. 1935 年，Hasse 和 Davenport 对于椭圆曲线证明了此猜想. 40年代初期，Weil 采用数论中的 Gauss 和与 Jacobi 和对于 Fermat 曲线 $X^n + Y^n = 1$ 和 Artin-Schreier 曲线证明了此猜想，并且正是基于这些例子，Weil 才提出这个结果对于任意曲线均应成立. 这个猜想首先是由 Weil 本人于 1948 年证明的. 为了证明这个猜想，他写了一本书《代数几何基础》，这本书客观上大大推动了代数几何学的发展. 1971 年，苏联数学家 Степанов 给了一个初等证明，后来又由 W. Schmidt 所简化. 而至今最简单的初等证明是由 Bombieri 于 1976 年给出的，只用到 Riemann-Roch 定理.

Weil 还研究了 Grassman 簇，Abel 簇等高维代数簇，在此基础上提出了如下的高维猜想：设 V 是有限域 $\mathbf{F}_q$ 上的非奇异绝对不可

约 r 维射影代数簇，N_n 为 V 在 F_{q^n} 中的射影点数. $Z_V(T)=\exp\left(\sum_{n\geqslant 1}N_nT^n/n\right)$ 为 V 的 zeta 函数. 则

(1)(有理性)$Z_V(T)=\dfrac{P_1(T)P_3(T)\cdots P_{2r-1}(T)}{P_0(T)P_2(T)\cdots P_{2r}(T)}$,

其中 $P_h(T)\in \mathbf{Z}[T](0\leqslant h\leqslant 2r)$, $P_0(T)=1-T$, $P_{2r}(T)=1-g^rT$;

(2)(Riemann 猜想的模拟) 设 $\deg P_h=B_h$(B_h 叫 Betti 数，它是某个上同调群的维数)，则

$$P_h(T)=\prod_{i=1}^{B_h}(1-\alpha_{hi}T)\quad(0\leqslant h\leqslant 2r),$$

其中 $|\alpha_{hi}|=q^{h/2}$;

(3)(函数方程) $Z_v\left(\dfrac{1}{q^rT}\right)=\pm q^{\frac{rk}{2}}T^kZ_V(T)$,

其中 $k=\sum_{h=0}^{2r}(-1)^hB_h$;

(4)(解数估计) $N_n=\sum_{h=0}^{2r}(-1)^h\sum_{i=1}^{B_h}\alpha_{hi}^n=q^r+O\left(q^{r-\frac{1}{2}}\right)$,

其中 O 项与 n 无关.

1960年，Dwork用p进分析工具证明了$Z_V(T)$的有理性. 1964年，M. Artin(老E. Artin的儿子)和法国数学家Grothendieck用后者发展的复杂的代数几何工具证明了该猜想的大部分内容. 剩下最困难的 $|\alpha_{hi}|=q^{h/2}$ 则是由 P. Deligne 于 1973 年证明的. Grothendieck 在代数几何方面的工作和 Deligne 对 Weil 猜想的最终证明分别得到 1966 和 1978 年的 Fields 奖.

有限域理论在 20 世纪中期用于构作各种试验设计方案. 随后便有效地应用到计算机科学和通讯工程上，有限域上代数曲线的 Weil 定理用来构作相关性能良好的二元序列或纠错性能良好的线性码. 最近由苏联数学家 Goppa 等人开创的代数几何码，其纠

错性能优于前人构作的码,为工程技术开辟了广泛的远景和理论指导作用(读者不妨重读一下 §1 中所引的美国数学家 1983 年报告中那段话).

最后我们扼要谈谈有限域上代数曲线算术理论与代数数论是如何统一起来的. 对于有限域 F_q 上的代数曲线 $C: f(x,y) = 0$, $f(x,y) \in F_q[x,y]$,则域 $F_q(x,y)$ 叫作曲线 C 的有理函数域,它是域 $k = F_q(x)$ 的有限(代数)扩张,因为 x 和 y 满足 $f(x,y) = 0$. 反之,k 的每个有限可分扩张均有形式 $K = k(y) = F_q(x,y)$,其中 y 在 k 上代数,即满足某个方程 $C: f(x,y) = 0$,即 K 是曲线 C 的有理函数域. 有理函数域 $k = F_q(x)$ 有许多与有理数域 Q 相似的性质. 例如 Q 是整数环 Z 的商域,k 是多项式环 $F_q[x]$ 的商域,而 Z 和 $F_q[x]$ 均是主理想整环. k 的有限扩张 K 叫作系数为有限域的单变量代数函数域,今后简称函数域. O_K 是 K 中对 $F_q[x]$ 整的元素全体,这也是 Dedekind 整环,叫作 K 的整数环. 从而函数域 K 具有和代数数域很相似的各种整体性质和理论(如素理想分解法则,分歧理论,类域论等). 另一方面,函数域也有与代数数域相近的赋值理论,从而有相近的局部域理论. 现代数论将代数数域与函数域统一起来,统称整体域(Global Fields),而它们对某个赋值的局部化叫作局部域(Local Fields). 许多数学理论和论断可以统一在所有整体域上加以考虑. 例如:Faltings 结果对于函数域情形于 1963 年就由 Manin 证明了. 又如:设 K 是有限域上曲线 C 的有理函数域. 我们在前面已经定义了曲线 C 的 zeta 函数 $\zeta_C(s) = Z_C(q^{-s})$,我们还可像代数数域那样定义函数域 K 的 Dedekind zeta 函数 $\zeta_K(s) = \sum_A N(A)^{-s}$(其中 A 过 O_K 的非零整理想). 可以证明事实上 $\zeta_K(s) = \zeta_C(s)$!(从而函数域 K 的“Riemann 猜想”是由 Weil 于 1948 年证明了.)再如:我们于 §6 中介绍过,猜想 Q 上椭圆曲线可以有任意大的秩. 而对函数域的情形,Шафаревич(苏联)和 J.

Tate(美)于1967年就证明了：$F_q(x)$上椭圆曲线可以有任意大的秩. 当然这并不是说，函数域上的事情总比代数数域上简单. 比如说，Kronecker-Weber在19世纪证明了：Q的最大Abel扩张Q^{ab}是所有分圆域的合成. 而直到1974年，Hayes才用类域论具体构作出$F_q(x)$的最大Abel扩域来，它是一些“分圆函数域”的合成，而这些分圆函数域的构造比分圆数域$Q(\zeta_n)$复杂. 分圆数域已发展成为内容十分丰富的一门学科，而分圆函数域的理论相当于秩1的Drinfeld模理论，是80年代才蓬勃展开的，有许多未解决的问题.

§8 互反律和Langlands纲领

> 在解决一个数学问题时，如果我们没有获得成功，原因常常在于我们没有认识到更一般的观点，即眼下要解决的问题不过是一连串有关问题中的一个环节.
>
> ——Hilbert:《数学问题》，1900年.

现在我们介绍Hilbert第9问题. 这个问题是要求在任意代数数域中建立像Q中Gauss二次互反律那样的一般互反律. 所谓Gauss二次互反律是指：对于任意两个不同的奇素数p和q，

$$\left(\frac{q}{p}\right)=\left(\frac{p}{q}\right)(-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}=\left(\frac{p^*}{q}\right),$$

其中$p^*=(-1)^{\frac{p-1}{2}}p$，而$\left(\frac{p}{q}\right)$为Legendre符号，即对于$p\mapsto a$，

$$\left(\frac{a}{p}\right)=\begin{cases}1, & \text{若 } a \text{ 是模 } p \text{ 的二次剩余；}\\ -1 & \text{否则.}\end{cases}$$

人们早就考虑Gauss二次互反律的推广工作. 对于$l\geqslant 3$，在19世纪发现了一些特殊的l次幂互反律，另一方面，人们发现了Gauss二次互反律的许多种证法，比如说，可以用分圆域中素数分解律来证明Gauss二次互反律，这导致Hilbert对于Gauss互反律给予新

的解释，他对于任意代数数域K定义了Hilbert符号$\left(\frac{a,b}{P}\right)$（其中$P$对应$K$的某个赋值），定义方法与§5介绍的（$K=Q$）完全一样. 他证明了$K$上的Hilbert符号的乘积公式，将它看成是域$K$上的二次互反律. 为了将$l$次幂互反律推广到任意代数数域上，Hilbert认为需要研究数域的Abel扩张. 而1920年高木贞治的类域论正是作了这件事情. 1927年，E. Artin对于数域的任意有限Galois扩张L/K构作了一个L-函数$L(s,L/K,\chi)$，其中χ是Galois群$\mathrm{Gal}(L/K)$的一个表示的特征. 当L/K是Abel扩张时，它就是通常的Abel$L$-函数. 为了证明这一点，他证明了Abel群$\mathrm{Gal}(L/K)$同构于K的广义理想群的一个商群. Artin把这个同构看成是一般互反律. 后来，Hasse利用类域论对于任意l定义了任意数域K上的l次幂剩余符号，建立了l次幂互反律. 这样，Hibert第9问题的最基本任务便完成了.

40年代以后，互反律的推广主要有三个方向. 一个是Artin互反律推广到高维上同调群上（类域论只涉及到二维上同调群），J. Tate等人对于无限Galois扩张得到一般对偶定理. 这些Galois上同调于70年代又被M. Artin和B. Mazur用étale上同调重新加以解释和推广. 一是60年代的代数K-理论对于l次幂剩余符号和范剩余符号赋以新的意义. 而最大的推广则是将关于Abel扩张的Artin互反律推广到数域非Abel扩张L/K中去. 或者从解析的观点，即是对于（非Abel的）Galois扩张L/K，研究Artin$L$-函数$L(s,L/K,\chi)$. Artin证明了它满足一个函数方程，40年代初期R. Brauer将它解析开拓成整个复平面上的亚纯函数. 而Artin猜想它除了在$s=1$处为极点之外，在其余地方均是解析的.

1967年，美国数学家Langlands研究了一般的约化代数群上自守形式的解析理论，看到Artin$L$-函数和Eisenstein级数理论中产生的Euler乘积之间形式上的联系，这使他提出非常一般的猜

想. 下面是这个猜想的最简单情形：设 L/Q 是 Galois 扩张，R：$\mathrm{Gal}(L/Q) \to \mathrm{GL}(2,C)$ 为 Galois 群 $\mathrm{Gal}(L/Q)$ 的一个 2 次不可约表示，χ 是以某种方式由表示 R 给出的 Dirichlet 特征：$L(s,R) = L(s,L/K,\chi) = \sum_{n\geq 1} a_n n^{-s}$ 是 Artin$L$- 函数，Langlands 猜想 $f_R(z) = \sum_{n\geq 1} a_n q^n (q = e^{2\pi iz})$ 是权为 1 的某种"新"的模形式. 且猜想每个权 1 的新模形式均可用此种方式得到. 1969 年，Langlands 证明了：$f_R(z)$ 是新的模形式 $\Leftrightarrow$ 对于 $\mathrm{Gal}(L/K)$ 的每个一次表示 R_1，$L(s, RR_1)$ 均满足 Artin 猜想. 1973 年，J-P. Serre 和 P. Deligne 证明了上述 Langlands 猜想.

Langlands 的一般猜想是关于任意整体域 Galois 扩张 L/K 的任意 n 次 Galois 表示 R：$\mathrm{Gal}(L/K) \to \mathrm{GL}(n,C)$ 的论断. 甚至于更一般的是关于任意约化代数群（不只是 $\mathrm{GL}(n,C)$ 以及由代数簇上同调群给出的 $\mathrm{Gal}(L/K)$ 的 p 进表示的论断. 我们不想详细叙述 Langlands 一般猜想的确切内容和意义. 只想再指出，当 $n = 1$ 时，Langlands 猜想相当于 Artin 互反律，或者说相当于关于整体域 Abel 扩张的类域论. 从而 Langlands 猜想也可以看成是类域论的高维（即对非 Abel 情形）的推广. 再想指出：§ 7 中所述的关于椭圆曲线 L- 函数与权 2 模形式之间联系的谷山 -Weil 猜想也是 Langlands 猜想的一个特殊情形.

Langlands 猜想是如此庞大，以至于在表面上一下子很难再认出古典代数数论的影子. 它提供了代数数论更深刻的背景，指出数论与代数几何，代数群的表示理论，自守形式（模形式的推广）理论，Abel 簇（椭圆曲线的高维推广）理论等许多现代数学分支之间的深刻联系. 人们把这整套构思称作 Langlands 纲领或 Langlnads philosophy. 希望有志青年去学习这套诱人的理论，攻克其中的一道道难关，攀登现代数学的高峰！[后记 4]

后　　记

1. 近来已经证明了猜想 1 对于每个二次数域 K 是正确的，并且决定出 K 上所有椭圆曲线的 K-torsion 子群的（有限个）结构.

2. 1993 年 6 月 23 日上午 10 时 30 分，40 岁的英国数学家 A. Wiles 在英国剑桥大学牛顿数学所做完题为“椭圆曲线，模形式和 Galois 表示”的演讲之后，宣布他对于半稳定（semistable）的椭圆曲线证明了谷山-Weil 猜想，并由此证明了 Fermat 猜想，最终解决了这个具有三百余年历史的著名数学问题.

3. K. Rubin 于 1989 年证明了：对于具有复乘的椭圆曲线，BSD 猜想的奇数部分是正确的. 本文作者对于一些具有复乘的椭圆曲线，验证了 BSD 猜想的偶数部分也是正确的，从而对于这些椭圆曲线，整个 BSD 猜想均正确.

4. 前苏联数学家 Drinfeld 证明了函数域上二维情形的局部 Langlands 猜想. 这项工作以及在量子群等方面的工作，使他得到 1990 年菲尔兹数学奖.

—— 补记于 1993 年 9 月

1

Ⅷ

最优化、辨识与控制

Optimization, Identification and Control

越民义
（中国科学院应用数学研究所）
Yue Mingyi
（Institute of Applied Mathematics，
Academia Sinica）

最优化介绍
An Introduction to Optimization

中国数学发展的若干主攻方向
Some Main Aspects on Development
of Mathematics in China

人们要达到某一目的，常会受到许多限制(约束)，例如人力、物力、财力、环境，等等. 在所给定的约束之下，要达到所期目的，往往有若干条途径可循，在这些途径当中如何选取一条，使得某种(或某几种)指标达到最优，例如花费最少，耗能最低，时间最短，等等，这就是最优化这门学科所研究的对象. 这里所说的最优化实际上指最优化**方法**. 所考虑的主要是针对不同性质的情况(例如约束由线性(不)等式组和(或)非线性不等式所定义，某些变量只能取非负整数值、某些数量带有随机性，等等)，设计出寻求使某种给定的指标达到最优值的方法. 因此，最优化这门学科的内容主要由三个部分组成：方法的设计、理论上的论证、计算机上的实现.

最优控制系统理论也可以认为是一种最优化理论. 但最优控制理论所研究的是要对某一给定的控制系统寻求一控制函数，在它的作用下，系统从一个状态以某种"最佳的"方式转移到另一状态. 它与我们上面所说的最优化是不相同的.

最优化这门学科，根据研究方法的不同，分成两个大的分支：数学规划和组合优化. 前者主要处理变量为连续型的问题，后者则处理离散性的问题. 在国际会议上，常把组合优化列入数学规划的一个分支. 但由于两者在问题的提法和使用的方法上大异其趣，我们便把它们分开来叙述.

最优化一词容易使人想起微分学中的求极值问题算法. 但这两者极少有共同之处. 经典的求极值算法所考虑的主要是可微分的函数. 求极值则是通过解一组联立方程得到. 因而它只适用于一些变量个数少的，函数结构简单的问题. 而生产实际中所出现的问题，正如前面所说，可能包含很多个变量，例如几千、几万、几十万、……. 问题中所出现的函数可能很复杂. 变量可能只限于在某类集合中取值，等等. 这些问题在计算机科学未出现之前，人们是不会去考虑的.

不可能存在一种最优化算法可以用来对付所有的问题. 因此，

最优化这门学科的研究内容可以说是无止境的.即使像线性规划这样看来是很简单的问题,从它的出现到今天仍不断有人在研究.一个算法的好坏一般取决于三个因素,即适应面要广,从理论上能严格证明:根据这一算法能够得到所要的解或其某种近似;其次,要求这一算法在计算机上容易实现;此外,还要求有较快的收敛速度.

在数学上,根据用来处理问题的方法,人们将数学规划分成若干分支.首先是将问题分成决定性的和随机性的两种.所谓决定性的,是指问题中所出现的参数都有确定的值或明确的取值范围,所有的变量皆非随机变量.反之,若问题中某些参数只能从某种统计意义上知道,或其变量中有随机变量,则称之为随机规划问题.下面分头作些介绍.

数学规划(决定性的) 早期的数学规划问题可以简单写成

$$(P) \qquad \min\{f(x)\mid x\in\Omega\},$$

这里的 Ω 是 n 维欧氏空间中一个给定的集合,$f(x)$ 是一个给定的函数.若 $f(x)=C^Tx$,C 为一 n 维常数向量,Ω 为由若干个线性(不)等式所限制而成的集,则称(P)为一线性规划.否则根据不同的情况便有不同的名称.例如当 Ω 是由若干(不)等式所定,若这些(不)等式或(和)$f(x)$ 中出现非线性函数,则称之为非线性规划;若 Ω 是由某些整点所组成,则称之为整数规划;等等.

线性规划的用途是在1947年Dantzig的单纯形法出现之后才受到广泛注意的.Fourier在1823年,de la Vallee Poussin在1911年和Kantorovich在1939年皆已注意到线性规划在实际问题中应用的可能性.尤其是Kantorovich的工作,它已指出不少实际问题皆可形成线性规划来求解.但他们的工作并未引起人们广泛的注意.这可能是由于没有提出一个较有效的解法以及快速计算技术尚未得到发展的缘故.关于线性规划的研究,目前主要集中于关于大型问题的解法上面,即当变量个数上万、几十万、……时,如

何设计出比现存的更快的算法. 在这方面，主要的工作是关于分解算法的改进，关于并行算法的研究，等等. 内点法在近十年来已成为有关线性规划问题研究的一个热点，虽然对它的成效有着各种相互矛盾的传闻. 关于分解算法，目前最大的测试问题的规模已达二百万个可能方案(Scenario)和十亿个约束.

线性规划的一个自然的延伸就是非线性规划，以下略写作NLP. NLP虽然是线性规划的对立之词，但今天所说的NLP却有其固定的含义，即限于在(P)中的$f(x)$和(或)用来定义Ω的(不)等式中出现非线性函数的情况. 就是说，我们将诸如整数规划、参数规划、随机规划等都另立户头，不纳入NLP.

NLP的研究始于1950年Kuhn和Tucker的文章，其中主要论证了在一定条件下NLP的解必须满足的一组条件，通称K-T条件. 文中并指出，对于凸规划来说(即$f(x)$为一凸函数，Ω为一凸集)，古典的Lagrange乘子法可以适当地推广到具有不等式约束的情形. 关于极值条件，下面还将谈到. 自Kuhn和Tucker的文章出现之后，有关NLP文章有如雨后春笋，大量出现，其内容大致可以分为四个方面：

1. 算法 即用来求出所与问题的解的方法. 不言而喻，这是数学规划的核心部分. 早期的算法多为直接法，即根据某种直观的判断设计出来的方法. 如所谓"登山法"、"翻筋斗法"之类. 这类方法一般多用于对解的精确度要求不高的情形，而且多用于无约束规划(即$\Omega = R^n$)；或目标函数的表达式未知，它在某一点的值只能依靠作实验等方法得出；或目标函数的变化比较平和等情况. 现在人们所研究的主要是一些迭代法. 对于设计出来的一个方法，我们须从理论上论证：根据所给的方法，在有关函数满足必要的条件(这是伴随方法给出的)之下，我们能够得到所要的最优解(一般是达到所需的精确度的近似解). 这就是算法的收敛性. 一个算法的好坏取决于(Ⅰ)适应面的广度，即有关假设越弱越好. (Ⅱ)在

计算机上易于处理,和(Ⅱ)收敛速度,即对于求出所需的精确度的解所必须花费的运算次数(或时间).由于没有一种算法能适用于所有的实际情况,关于算法的研究可以说是无止境的.我国的运筹学工作者在这方面作了不少很有意义的工作.

2. 有关极值点的条件 即用来判断所求出的点是否为所求的解的条件.本节开始所说的 K-T 条件即为一种必要条件;对于凸规划来说,它也是充分的.类似的还有 F. John 条件、Slater 条件等.这一类条件只适用问题所涉及的函数都是可微的情形.关于不可微规划,情况要复杂得多.

对每一个非直接算法,我们都要讨论它的收敛性和通过它所求出的值是否满足极值条件.由于现存的方法很多,人们不禁要问,能否从某种统一的观点来研究算法.就是说,我们可以将具有某种共性的算法放在一起,抽象出它们的共同性质,设计出一种具有这些共性的抽象算法,从而讨论其收敛性和极值条件.在这方面,目前已有两条途径可循.一条是从点到集映象来研究,即将一次迭代看成是一个点到集的映象,由此来讨论算法的性质.这一途径是在 1969 年由 Zangwill 开始的,其后得到了进一步的发展(可参考作者在 1982 年的《数学年刊》第四期上发表的文章).另一条途径是引入参数,即设计一种带参数的算法.当赋与这些参数以某种具体数值之后,便可得到一个具体的算法.在这方面,对于无约束规划,是由 H. Y. Huang 在 1969 年开始的.对于带约束的情形则是作者与韩继业教授在 1982 年(《应用数学学报》)开始的.这方面的工作还有待进一步的发展.比如说,参数本身对于算法的基本性质的影响如何,如何选取参数可以得到具有某种所需要的性质的算法等问题仍有待回答.

3. 对偶性 我们称(P)为原问题.从(P)我们可以定义一个与之伴随的问题(P^*):

$$(P^*) \qquad \max\{g(y) \mid y \in \pi\},$$

称为(P)的对偶. (P^*)须要具备三个性质,即(Ⅰ)若(P)是求$f(x)$的极小(大)值,则(P^*)是求$g(y)$的极大(小)值;(Ⅱ)常有$f(x) \geqslant g(y), \forall\ x \in \Omega, y \in \pi$;(Ⅲ)若$x^*$和$y^*$分别为$(P)$和$(P^*)$的解,则有$f(x^*) = g(y^*)$. 研究对偶问题的主要目的有三:(Ⅰ)可以利用$(P^*)$来帮助求$(P)$的解;(Ⅱ)可以帮助判定$(P)$是否有解;(Ⅲ)可以帮助解释$(P)$的解的实际意义,特别是有关经济方面的问题. 对偶问题的研究是在线性规划出现不久便开始的,当Dantzig带着他发现的单纯形法去与J. von Neumann讨论时,他从von Neumann那里获知这一概念(当时von Neumann正与Morgenstein完成他们的Theory of Game一书). 这一方向已得到相当充分的发展,可参考Rockafellar的

The theory of subgradients and its applications to problems of optimization: convex and nonconvex functions (1981)一书.

4. 稳定性 这是指当问题的某些参数发生微小变化时,问题的解答会发生什么样的变化. 在生产实际中,与问题有关的某些数值,如价格,时间,等等,不可能长期稳定不变. 它们的微小变化有时会给问题的解答带来大的变化. 假若能知道参数在什么样的范围内变化,问题的解所出现的变化是可以容忍的,则可省去大量的计算费用和时间.

稳定性问题的研究是参数规划的一个组成部分. 参数规划的一般形式为

$$OP(z) = \min\{f(x,z) \mid x \in \Omega(z)\}, z \in Z.$$

这种写法已成通用. 但它远不能反映参数规划所研究的内容. 它仅表示:这是一簇数学规划. 参数规划的内容除了稳定性之外,还包括:求出集$B = \{z \in Z \mid \Omega(z) \neq \varnothing\}$,求出集$A = \{z \in Z \mid OP(z)$可解$\}$,以及求$z \in Z$,使$OP(z)$的解集具有事先要求的某种结构,等等. 参数规划的一个解是指求出一个$Y \subseteq Z$,当$z \in Y$时,相应的$OP(z)$的解对于某种事先给定的性质是不变的. 参数规划已成为

一门内容丰富，研究相当深入的学科，可参考 K. Lommatzsch 所著：Anwendungender Linearen Parametrischen Optimierung, 1979.

决定性的数学规划还包括诸如动态规划，多目标规划，等等. 有兴趣的读者可参考《中国大百科全书》数学卷中有关的条目. 下面，我们将介绍近年来发展很快的一个分支：组合最优化.

组合最优化（或称**组合规划，离散优化**）. 组合最优化(CO) 是指关于(部分) 变量只取某些离散数值的优化问题的研究. 设有一离散的(一般是有限的) 集合 F. 设 $\mathscr{F}$ 是由 F 的某些子集作成的集，即若 $E \in \mathscr{F}$，则 $E \subseteq F$. 又设 $f(x)$ 为定义在 $\mathscr{F}$ 上的一实值函数，$f: \mathscr{F} \to R$，对于给定的 $\mathscr{F}$ 和 f，我们要求从 $\mathscr{F}$ 的元素中找出一个 E_0，使得 $f(E_0) = \min\{f(E) \mid E \in \mathscr{F}\}$. 若 $f(E) = \sum_{e \in E} f(e)$，$\forall E \in \mathscr{F}$，则称所与问题为线性的，否则称为非线性的.

在生产实际中存在着大量的 CO 问题. 比如厂房的选址，管道的铺设，工作日程表的制订，流水作业线的设计，人员的分派，计算机硬件的设计，编排、生产过程，等等，皆属于研究范围.

就实际中出现的绝大多数组合优化问题来说，上面提到的 $\mathscr{F}$ 是一个有限集. 因此最优解总是存在的. 人们也许会想到用列举法去求解. 但是生产实际中遇到的问题，$\mathscr{F}$ 的元素个数常常是天文数字，列举法一般是无效的. 对于数学规划来说，我们总是根据所与函数的性质(可微，不可微，凸，非凸，等等) 设计某些迭代性. 这使我们想到，可否根据 $f(x)$ 和 $\mathscr{F}$ 的某些性质设计出某种算法. 在这方面，人们曾作了大量的工作，例如关于拟阵、广拟阵、次模函数等的研究. 但由于实际问题中所出现的 $f(E)$ 和 $\mathscr{F}$ 的多样性，抽象出来的方法所能够解决的问题相对来说为数极少. 目前人们仍是针对一个个“典型问题”进行研究，去寻求某种“可行的”算法.

所谓典型问题是指：将性质相近的问题归入一类，从中取出一

个易为人们所了解的问题作为典型.例如"装箱问题"是一个典型问题:设有许多个容量相同(设为1)的箱子,又设有 n 件货物 J_1, $\cdots$, J_n 要装到这些箱子中去.设 J_i 的重量为 w_i: $0 < w_i \leqslant 1$,问题是如何装法可以用最少数的箱子将 J_1, $\cdots$, J_n 全部装进去.在实际中,这问题是常遇到的.例如钢铁厂中要用一定规格(比如10米长)的钢条去剪裁出用户提出的若干种不同尺寸的钢条.问:如何剪裁使得厂方能以最少数量的钢条剪裁出所要的钢条.类似的问题在造纸厂中也会出现.这些问题在实质上都是装箱问题.目前,著名的典型问题已在30个以上,其中某些又包含许许多多的子问题,例如时间表问题、旅行商问题等便是.著名的典型问题还有网络流问题、集的覆盖问题、分配问题、匹配问题、背包问题,等等.每个问题都有大量的实际背景和许多人在研究.

前面提到,对于最优化的问题来说,研究工作的主要目的是要寻求一种对所与问题的有效算法.什么是有效算法?可以装箱问题为例来说明.显而易见,对于所给的 n 个物件 J_1, $\cdots$, J_n, J_i 的重量为 w_i,总存在一个排列,按照这一排列装箱(每箱子所包含的物件的总重量不超过1),所用的箱子数量最少.要求出这一排列,若用列举法,则须将 $n!$ 个排列进行比较.当 $n = 20$ 时,$n!$ 已是天文数字了,便何况 n 为一千、一万.因此,列举法是无效的.当这一问题的参数 $n, w_1, \cdots, w_n$ 给定时,我们得到这问题的一"实例".将一实例在计算机上以某种计算机语言表出所需的长度 N 称为这问题实例的规模.若有一种算法,利用它,对于任何实例都可以在 αN^β 次运算之下将此实例的解求出(这里的 α 和 β 与 $n, w_1, \cdots, w_n$ 无关),则称此算法是有效的,这样的算法一般称为多项式算法.对于别的组合优化问题,"有效算法"可类似定义.不幸的是,已经证明:对于绝大多数著名的组合优化问题来说,若其中有一个有有效算法,则其他亦然.而对这些问题,至今尚未证实有一个有有效算法.这一否定性的结论可以认为是从抽象理论来研究组合问题的

一大贡献. 属于这类问题中的问题我们称之为NP完备问题. 当一个问题已被证明是NP完备问题，人们一般便不再去对它寻找有效解法了. 现已有成千的问题已被证明是NP完备问题. 关于早先的结果，读者可参考M. R. Garey和D. S. Johnson所著Computers and Intractability — A Guide to the Theory of NP-Completeness, 1979.

对于一个问题，即使已证明它是NP-完备的，我们仍有许多事情可做. 首先，若该问题的参数具有某种特殊结构，它可以是可解的，即在这种特殊结构的假设下，可能存在有效算法. 问题是如何去发现这些特殊结构. 例如旅行商问题. 这问题是这样的：设有 n 个城市，任何两城市之间皆有路连通，其距离为已知. 某旅行商从其中某城市出发，要求经过每城市一次，而且只是一次，然后回到原处. 要求他所走的路（总长）为最短. 这是一个著名的NP完备问题. 但若城市的距离 d_{ij} 之间存在某些关系，则问题是可解的. 可参考我国的《运筹学杂志》八卷二期(1989)上R. E. Burkard所著的"旅行商问题的一些特殊情况和启发式算法"一文.

其次，我们可以考虑近似算法. 所谓近似算法是指某些对于所给的问题简便易行的算法. 这里的问题是要求估计出这一算法好坏的程度. 设 P 为所给的问题，A 为一近似算法，I 为 P 的一实例，f 为目标函数. 对于 I 通过 A 得到 f 的值为 $f_A(I)$，设 $f_{opt}(I)$ 为 I 的最优解，A 的好坏是通过下述的最小的 $\alpha > 0$ 来衡量（假定 P 是求最小值）：

$$\frac{f_A(I)}{f_{opt}(I)} - 1 \leqslant \alpha,$$

式中 α 与 I 无关. α 越小，A 就越好. 若 $\alpha = 0$，则 A 为最优算法. 它保证，即使在最坏的情况之下，$f_A(I)$ 与 $f_{opt}(I)$ 之差至多为 $\alpha f_{opt}(I)$. 由于 α 的这一性质，近似算法的研究也称为最坏情况的研究.

以装箱问题为例，这是一个 NP 完备问题. 我们先将上述的 $w_i(i=1,\cdots,n)$ 按单降排列：$w_1 \geqslant w_2 \geqslant \cdots \geqslant w_n$，然后考虑算法 NFD. 它是将物件 $J_1, J_2, \cdots, J_n$ 按 $1, 2, \cdots, n$ 次序装箱，先装 J_1，若 J_2 能与 J_1 同装一个箱子，就把它们放在一起，否则即启用一个新的箱子. 一般说来，设 J_i 已经装好，若 J_{i+1} 能与 J_i 装入一个箱子，便将它们放在一起，否则即启用一个新的箱子. 可以证明 (Baker, B. S. and E. G. Coffman, 1981) 最小的 $\alpha = \alpha_{NFD} = 0.691\cdots$. 若当 J_i 已装好之后，在处理 J_{i+1} 时，我们将 J_{i+1} 按已装入物件的箱子的原有顺序逐一试装，若其中有能容纳 J_{i+1} 的箱子，则放入这种箱子中的第一个，否则启用新的箱子. 这种装法通称 FFD(first-fit-decreasing). 可以证明 (D. S. Johnson, 1973; B. S. Baker, 1985; Yue Minyi, 应用数学学报, 1992) 最小的 $\alpha = \alpha_{FFD} = \frac{2}{9} = 0.\dot{2}$. FFD 显然比 NFD 好得多. 若问题的提法是：设 $f_{OPT}(I) = m$，试问，在将箱子放大多少倍时，使得在FFD装法之下，m 只箱子可以将 I 装进去. 设最小的倍数为 α，可以证明 (M. Yue, Annals of OR, ed. by Hammer, 1989) $\alpha = \frac{13}{11}$.

由于大量的组合优化问题都是 NP 完备问题，近似算法可以认为是一个值得注意的发展方向. 遗憾的是：在这方面成功的例子并不多. 原因是任何有意义的改进都要高度的技巧.

第三条发展途径是：设计某些算法，它们虽然不是多项式的，但当问题的规模不太大时却很奏效. 比如对于旅行商问题. 在 70 年代，利用分支定界法，使用当时最先进的计算机，可以处理 40 个左右城市的问题. 而现在，通过关于多面体理论的研究，则可以处理成千上万个城市的问题. 考虑到这牵涉 1000! 种可能情况的选择，不能不说是一大进步.

生产实际中出现的组合优化问题常常是错综复杂的，不可能把它归结为上述的这一个或那一个典型问题. 但它的某些子问题

却可以利用有关典型问题研究所得到的成果去解决，因而整个问题可以综合利用某些方法去解决. 例如关于计算机的集成电路、模块和板块(Chips, modules and boards)的设计、布局与生产，所出现的问题则可综合利用旅行商问题、中国邮路问题、匹配问题、最大集团问题、图的分割问题、装箱与背包问题、最短路问题、边(或点)不连通问题、多种物资流问题、平面图上的Steiner树问题等以及它们的变种去解决. 关于这方面的情况，可参考 B. Korte, Applications of Combinatorial Optimization, 1988.

第四条发展途径是计算机模拟. 有些实际问题涉及的因素很多，各因素之间又存在着复杂的关系，要用数学方式去表述这类问题已是非常困难，更谈不上用数学方法去解决了. 针对所给定的实际问题，利用模拟技术，往往可以得到某种令人满意的解答. 关于这方面，读者可参考《运筹学杂志》在 1984 年所发表的 H. Müller-Merbach 的文章.

最后，我们要谈一下离散问题的某些特点. 一个组合优化问题的难易一方面取决于问题本身的提法，另一方面也取决于有关参数的结构. 例如旅行商问题是著名的 NP 完备问题，但中国邮路问题，即邮递员从某点出发，要将每条街至少走过一次回到原处，则是一个非 NP 完备问题. 关于同顺序的排序问题，两台机器的问题是容易的，而三台机器的问题则是 NP 完备问题. 又如树状图上的最短路问题，当边长皆为 1 时是容易的，若容许边长可以是 1，也可以是 2，则成了难题. 因此，我们可以说，组合优化包含了无穷无尽的难于处理的问题. 它们皆可称得上真正的问题，使数学家大有用武之地. 对于其中每一问题的解决皆有可能给生产实际带来益处.

组合优化的应用是非常广泛的，生产管理中出现的许多优化问题皆属于组合优化问题. 一些乍看起来似乎与它不相干的问题只须经过适当的转换便可化成这类问题. 例如某些生产单位在作预算、投资等就出现这样的问题：设全年的资金总额为 A，有 n 个

项目可做. 但限于资金，不能全做，如何从中选择一部分使总的收益最大？设投资于项目中 i 须耗资 a_i 元，收益为 b_i 元，则问题即变成求 $x_i(x_i=0$ 或 $1)$，使得在条件 $\sum_{i=1}^{n}a_ix_i\leqslant A$ 之下，$\sum_{i=1}^{n}b_ix_i$ 达到最大，这就是上面提到的背包问题的一种特殊情况：0-1 背包问题. 如 x_i 不限于取 0 或 1，则得一般的背包问题. 又如某工厂要生产 n 种耐火材料 $m_1,\cdots,m_n$. 设对生产 m_i 的要求是：温度必须从 s_i 开始到 f_i 为止，设在生产 m_i 之后接着生产 m_j，则须将炉子的温度从 f_i 按情况降或升到 s_j. 设炉子从 b 度下降到 a 度 $(b>a)$ 的费用为 $\int_a^b g(x)\mathrm{d}x$，从 b 度升到 a 度 $(a>b)$ 的费用为 $\int_b^a f(x)\mathrm{d}x$. 试问如何将 $m_1,\cdots,m_n$ 安排一个顺序，使总的费用最省？这是一个排序问题，是旅行商问题的一个特例. 在军事方面也存在大量的例子，比如：(Ⅰ) 在攻击手段受到某种限制的条件下，如何确定一种最优阻止策略，以破坏敌方的通讯或交通网络；(Ⅱ) 确定一交通网络在受到敌方破坏性攻击的情况下的灵敏度，从而（根据某种准则）定出最优措施，以疏散某些保护流通的设施.

上述这些问题是早已存在的，只是由于受到计算技术的限制，现实意义不大，未能引起人们普遍的关注. 现在，由于生产组织管理的高度复杂化、市场竞争的国际化，以及产品更新换代的加速，使得大量的组合优化问题被提了出来. 由于计算技术的高度发展，使得有可能对它们提出解答或某种令人满意的解答，这就使得这门学科很快得到发展. 无怪像以研究测度论闻名的数学家 P. R. Halmos 也说(1981)：“在可以预见的将来（以及现在），在试图了解世界方面，离散数学将成为日益有用的工具，而分析则扮演一个相对来说次要的脚色.”

数学应该而且能够为人类的生产实际、为改善人类的生活提供某些帮助. 它不同于文学，一篇好的文学作品可以激发人们前

进，可以帮助认识人与人之间或人与社会之间的关系，可以帮助人去更深刻地了解种种社会现象．但一篇纯数学的文章往往只有那么少数几个搞同类问题的人才能欣赏．多数纯数学文章对于数学的发展、对于人类的前进无所裨益．著名的大数论学家 G. H. Hardy 在他的"A Mathematician's Apology"(1940) 一书中曾说："我曾帮助培养别的数学家，但不过是与我同样类型的数学家，他们的工作，至少就我所帮助他们的那部分来说，与我的工作一样是无用的．从任何实用的准则来看，我的数学生命的价值等于零．"著名的数学家 J. von Neumann 曾说过(1947)："当一门数学发展到远离它的经验的泉源，或更有甚者，假若它只是为来自于'实际'的某些概念间接地诱发出来的(一门学科的)第二代或第三代，它就要陷入严重的危险之中．它就变得越来越唯美主义，越来越为艺术而艺术．假若这一领域的周围还有一些相互关联、仍然与经验有密切联系的学科，或者是，这门学科是在某些具有非常健康的鉴别力的人物的影响之下，这不一定是坏事．但若这门学科是沿着一条阻力最小的道路前进，则这条渠道，非但不是泉源，而且还分成许多无意义的分支，这门学科就变成一堆散乱的枝节和繁琐的事物．"

上面引述了几段话，主要是用来反映某些著名的数学家对于数学这门学科的看法，可以供读者参考．

陈翰馥
(中国科学院系统科学研究所)
Chen Hanfu
(Institute of Systems Science,
Academia Sinica)

辨识与控制
Identification and Control

对控制的研究和应用可追溯到19世纪，甚至更早，但它的辉煌成就却是近30年来取得的．这和现代高技术的发展过程中提出一系列要求精确、快速的控制课题有关，并且执行控制使命的环境，有时是恶劣的，甚至是人所不能及的．例如用防空导弹来阻击来袭目标，既要快，又要准，还只能用遥控的办法，并且首先要用辨识和估计的办法去识别真假弹头，跟踪目标．Apollo（阿波罗）登月舱的导航控制，既要它不偏离预先设计的最优轨线，并且到月球时要软着陆．在航空方面，飞机的自动驾驶仪，为提高机动性能而设计的空气动力学不稳定飞机的飞行控制，为节省燃料飞机最优升、降轨线的设计等都是控制的成功应用．在工业过程控制方面，有不胜枚举的成功例子．例如钢铁工业中的温度控制，轧钢控制，化工流程控制，巨型油轮的适应控制，造纸工业中的控制，大型电力系统的控制等，不仅提高了产品或运行的质量，而且有很大的经济效益．特别涉及到要接触有害物资或困难环境下的操作，例如核反应堆控制，水底操作等，利用控制技术，不只提高经济效益，而且有效地提高了安全度．控制在技术领域中取得如此灿烂的成就和计算机的发展有密切关系，工程技术提出了快速、精确的要求，计算机就为这种快速、精确的控制的技术实现提供了可能．

工业发展特别是高技术发展对控制提出的迫切要求和计算机发展所提供的可能性，就对理论和方法提出了一系列既有应用需要又有理论深度的课题，吸引了一大批科学工作者，导致了控制理论的大发展．理论和方法上取得的重大进展是控制技术取得一系列引人瞩目的成就的重要因素．

反馈是控制理论中的一个首要概念，19世纪中叶考察蒸气机在调速阀作用下的稳定性就是对反馈控制的研究．所谓反馈就是用系统输出信号中所包含的信息来设计改善系统的输入（或控制）信号，使系统达到预定目标，例如要使系统的输出达到某个设定值．自动调节原理就是用反馈的办法来调节用常系数常微分方程

描述的系统，这是控制理论的经典部分. 我国著名科学家钱学森是这方面的著名专家，他的专著《工程控制论》在国际上享有盛誉.

标志着现代控制理论诞生，并且在它今后的发展起着深远影响的工作有以下几个方面：一是苏联 Понтрягин 的极大值原理，这指当系统由常微分方程描述，为使一个性能指标达到极值，控制作用应满足的必要条件. 由于控制系统中控制的取值范围一般是闭集，经典的 Lagrange 乘子法不适用，而极大值原理却可以处理. 二是美国 Bellman 的动态规划，他从"最优化准则"出发，即认为从最优轨线任何一个中间点到终点的轨线必是以这个中间点为初值的最优轨线，导出动态规划方程，解出这个方程就可以得到最优控制. 这两方面的工作都围绕着使性能指标达极值的最优控制问题. 但系统本身和量测到的有关系统的信息可能都伴有随机干扰，这就要用滤波的办法处理噪声. 在第二次世界大战期间，美国由于高炮技术精度的要求，Wiener 提出了信号为平稳过程的滤波方法. 在苏联 Колмогоров 也提出了类似的滤波方法，这后来称为 Wiener-Колмогоров 滤波. 这一套方法不仅在理论上，而且在应用中都起过很大作用. 随着计算机技术的发展，并结合对滤波精度提出更高要求，适时递推地处理不一定是平稳的信号，就有极大意义. 应运而生的就是 Kalman 滤波，这是现代控制理论形成的另一个重要标志. 利用 Kalman 滤波，对我们感兴趣的量的估计，可利用新量测到的数据，适时地加以修正，所以它在制导、导航及工业过程控制等领域，得到许多重要的成功应用. 标志现代控制理论诞生的另一组工作是 60 年代以来对用状态空间描述的线性控制系统能控性、能观测性、反馈镇定、输入输出系统的状态空间实现等概念的提出和研究. 能控性是指控制作用能否把系统的状态控制到零点. 能观测性是指对系统状态的部分观测能否定出系统的全部状态. 反馈镇定是指用反馈控制的办法，把原先不稳定的系统，在闭环的状态下成了稳定的系统，这方面的研究成果，在许多国家

都已做成了计算机辅助设计软件包，得到了许多成功的应用.

为了控制一个系统，首先要建立这个系统的数学模型.系统辨识就是根据系统的输入、输出数据来建立系统的数学模型.例如一个复杂的化工流程所得到的产品的数量和质量指标(输出)依赖于外加的催化剂、温度和压力(输入)等，但我们事先不确知它们的依赖关系.系统辨识就是根据化工过程的温度、压力和所加的催化剂以及产品的数量和质量指标来建立它们之间的动态关系.系统辨识分“离线”和“在线”两种.离线的系统辨识就是先积累系统的输入输出数据，然后利用积累起来的数据来建模，在实用中就把这样建立起来的模型当作真实系统来对待.但有相当多的一类实际问题，有些根本不可能事先辨识，例如在拦截问题中对来袭弹头气动系数的估计，只能当它来袭时逐步加以识别，有些则因为系统本身可能会有慢的或随机的变化，必须要在获得输入输出数据的同时，建立模型，随时修正模型，并根据这样不断修正的模型给出控制，这就是适应控制.

从客观的物理系统所量到的输入输出数据，可能用有限维的确定性微分方程或差分方程去拟合就能满足所要求的精度.但有时量测数据中的随机噪声或系统本身的随机因素不能忽略，这时就要用随机微分或差分方程去拟合，这就是随机系统的辨识.有些系统，例如大型柔性天线，加热炉的温度分布等，就要用偏微分方程去拟合，估计偏微分方程中未知的函数或参数就是分布参数系统辨识.但最便于计算机在线辨识和控制，并已取得许多成功应用的莫过于受外干扰的自回归滑动平均过程，即 ARMAX 过程作为模型来辨识.

ARMAX 模型就是系统的 l 维输入 u_n 及 m 维输出 y_n，由下述随机差分方程来描述

$$
\begin{aligned}
& y_{n+1} + A_1 y_n + \cdots + A_p y_{n-p+1} \\
& = B_1 u_n + B_2 u_{n-1} + \cdots + B_q u_{n-q+1} + \varepsilon_{n+1}, \qquad (1)
\end{aligned}
$$

$$\varepsilon_{n+1} \overset{\triangle}{=} w_{n+1} + C_1 w_n + \cdots + C_r w_{n-r+1}$$

表示系统的噪声. 如果 $\varepsilon_n \equiv 0$, 那么系统(1) 就成了确定性系统. $\{w_n\}$ 是不相关的驱动噪声,通常假定$\{w_n, \mathscr{F}_n\}$ 是一个鞅差序列,或是一串独立噪声序列.

系统辨识在 ARMAX 模型下,就是根据数据$\{u_i, i \leqslant n\}$, $\{y_i, i \leqslant n\}$ 给出系统阶数(p,q,r) 及参数阵

$$\theta^\tau \overset{\triangle}{=} [-A_1 \cdots - A_p B_1 \cdots B_q C_1 \cdots C_r] \tag{2}$$

的估计.

这个问题和时间序列分析中对 ARMAX 模型的估计有很大不同,因为模型(1) 中带有控制项 u_n,而它是依赖于$\{y_i, i \leqslant n\}$ 的反馈控制,所以过程$\{y_n\}$ 本身既非平稳也非遍历.

当 $r = 1$ 时,也就是 $\varepsilon_n \equiv w_n$ 时,如果系统阶数(p,q) 已经固定,那么估计 θ 一个最直观常用的办法是最小二乘.

记

$$\varphi_n^\tau = [y_n^\tau \cdots y_{n-p+1}^\tau u_n^\tau \cdots u_{n-q+1}^\tau], \tag{3}$$

那么系统(1) 可写成

$$y_{n+1} = \theta^\tau \varphi_n + w_{n+1}. \tag{4}$$

在 n 时刻对 θ 的最小二乘估计是指使下式达最小的 θ_n:

$$\sum_{i=1}^{n} (y_i - \theta_n^\tau \varphi_{i-1})(y_i - \theta_n^\tau \varphi_{i-1})^\tau$$

$$= \min_x \sum_{i=1}^{n} (y_i - x^\tau \varphi_{i-1})(y_i - x^\tau \varphi_{i-1})^\tau. \tag{5}$$

不难验算

$$\theta_n = P_n \sum_{i=0}^{n-1} \varphi_i y_{i+1}^\tau, \quad P_n = \left(\sum_{i=1}^{n-1} \varphi_i \varphi_i^\tau \right)^+. \tag{6}$$

若对 $n \geqslant n_0$, $\sum_{i=1}^{n-1} \varphi_i \varphi_i^\tau$成了非退化阵,那么 θ_n 可用递推的形式来表达:

$$\theta_{n+1} = \theta_n + a_n P_n \varphi_n (y_{n+1}^\tau - \varphi_n^\tau \theta_n), a_n = (1 + \varphi_n^\tau P_n \varphi_n)^{-1}, \quad (7)$$

$$P_{n+1} = P_n - a_n P_n \varphi_n \varphi_n^\tau P_n, \quad n \geqslant n_0. \quad (8)$$

从(4) 可以看出，

$$\hat{w}_{n+1} \stackrel{\triangle}{=} y_{n+1} - \theta_{n+1} \varphi_n \quad (9)$$

是对 w_{n+1} 的一个合理估计，所以对 $r \geqslant 1$ 的情形，把 φ_n 扩展成

$$\varphi_n^\tau = [y_n^\tau \cdots y_{n-p+1}^\tau u_n^\tau \cdots u_{n-q+1}^\tau \quad \hat{w}_n^\tau \cdots \quad \hat{w}_{n-r+1}^\tau]^\tau \quad (10)$$

后，在(7)，(8) 中任取初值 θ_0 及 $P_0 > 0$，那么(7)，(8) 仍可用来估计由(2) 表示的参数阵. 这时算法叫推广的最小二乘算法.

我们前面曾提到，实际系统的参数可能要随时间变化，但如果一个算法对定常参数都给不出好的估计，那它不能算一个好的算法. 所以我们自然要问，如果系统确由(1) 描述，最小二乘或推广的最小二乘算法是否具有强一致性，也就是当 $n \to \infty$ 时，θ_n 是否趋于 θ?

θ_n 的性质和阵 $\sum_{i=1}^{n-1} \varphi_i \varphi_i^\tau$ 的特性，特别和它的最大本征值 $\lambda_{\max}^n$ 及最小本征值 $\lambda_{\min}^n$ 有密切联系. $\limsup_{n\to\infty} \lambda_{\max}^n / \lambda_{\min}^n$ 有穷通常叫系统受到持续激励，这是经常用来保证 θ_n 收敛到 θ 的主要条件. 粗略地说，持续激励就是要求输入、输出信号不要过于简单，而要复杂到使系统的各种特性都激发出来. 比如说，如果输入信号恒为零，那么 $B_1 \cdots B_q$ 就无法估计出来，这时显见 $\lambda_{\min}^n$ 也恒为零. 所以为了辨识参数，某种激励是必需的，但是否一定要强到持续激励?

为了保证 θ_n 向 θ 收敛，用鞅估计为工具，可以发现，实际需要的条件比持续激励弱得多. 我们来叙述一个结果. 下面总记

$$A(z) = I + A_1 z + \cdots + A_p z^p,$$

$$B(z) = B_1 + B_2 z + \cdots + B_q z^{q-1},$$

$$C(z) = I + C_1 z + \cdots + C_r z^r.$$

定理 1 设$\{\mathscr{F}_n\}$ 是一族非降 σ- 代数，$\{w_n, \mathscr{F}_n\}$ 为鞅差序

列，$\sup_n E\{\|w_{n+1}\|^2/F_n\} < \infty$，并且

$$C^{-1}(e^{i\varphi}) + C^{-\tau}(e^{-i\varphi}) - I > 0, \forall \varphi \in [0, 2\pi] \qquad (11)$$

（这个条件叫严正实条件），那么

$$\|\theta - \theta_n\|^2 = O\left(\frac{\log\lambda_{\max}^n (\log\log\lambda_{\max}^n)^c}{\lambda_{\min}^n}\right), \forall c > 1. \qquad (12)$$

这个结果给出了 θ_n 对 θ 的估计误差阶数，要改进这个估计阶数，未必容易，但要减弱定理的条件还是有可能的. 例如可以把严正实条件(11) 减弱到只要求 $C(z)$ 稳定，但做了这样的减弱后，目前还写不出对 θ_n 的递推格式.

对 θ 的估计，也可以用别的算法. 一个最简易的算法是随机梯度算法：

$$\theta_{n+1} = \theta_n + \frac{\varphi_n}{r_n}(y_{n+1}^\tau - \varphi_n^\tau\theta_n), \theta_0 \text{ 任意}, \qquad (13)$$

$$\varphi_n^\tau = [y_n^\tau \cdots y_{n-p+1}^\tau u_n^\tau \cdots u_{n-q+1}^\tau \overline{w}_n^\tau \cdots \overline{w}_{n-r+1}^\tau], \qquad (14)$$

$$\overline{w}_n = y_n - \varphi_{n-1}^\tau\theta_{n-1}, r_n = 1 + \sum_{i=1}^n \|\varphi_i\|^2. \qquad (15)$$

和推广的最小二乘算法相类似，θ_n 的渐近性质和 $\sum_{i=1}^n \varphi_i\varphi_i^\tau$ 的最大、最小本征值 $\mu_{\max}^n$ 及 $\mu_{\min}^n$ 有密切关系.

定理 2 设$\{\mathscr{F}_n\}$是非降 σ- 代数族，$\{w_n, \mathscr{F}_n\}$ 是鞅差序列，$\sup_n E[\|w_{n+1}\|^2/\mathscr{F}_n] \triangleq \sigma^2 < \infty$, $E\sigma^2 < \infty$,

$$C(e^{i\varphi}) + C^\tau(e^{i\varphi}) - I > 0, \forall \varphi \in [0, 2\pi].$$

如果对某个 $a \in \left[0, \frac{1}{4}\right]$,

$$\mu_{\max}^n/\mu_{\min}^n \leqslant M(\log r_n)^{\frac{1}{4}-a}, \forall n \geqslant N, \limsup_{n\to\infty} r_n/r_{n-1} \leqslant M(\log r_n)^a,$$

并且 $r_n \to \infty$，那么由(13) 定义的 θ_n 当 $n \to \infty$ 时收敛到 θ.

上面两个定理说明，最大、最小本征值之间满足一定关系时，但不必持续激励，θ_n 就能收敛到 θ. 但什么样的输入，才能保证本征值之间的关系得到满足？

我们把要加在系统上的反馈控制记为 u_n^s，然后对 u_n^s 加上一个趋于零的扰动. 确切地说，对推广的最小二乘算法下的控制，我们加一个扰动 v_n，要求 $\{v_n\}$ 是相互独立的序列，并且与 $\{w_n\}$ 独立，$Ev_n = 0, Ev_n v_n^\tau = \frac{1}{n^\varepsilon} I, \|v_n\|^2 \leqslant \sigma^2/n^\varepsilon, \quad \forall\, \varepsilon \in \left(0, \frac{1}{2(t+1)}\right), t = \max(p,q,r) + mp - 1, \sigma^2$ 为常数，实际对系统的输入取为

$$u_n = u_n^s + v_n.$$

由于 $v_n \xrightarrow[n\to\infty]{} 0$，所以对诸如时间平均型的性能指标，$v_n$ 并不影响指标值. 但由于这个小激励，就可以得到对 θ 的一致估计.

定理3 设定理1的条件成立，还设 $A(z), B(z), C(z)$ 没有左公因子，A_p 满秩，$\frac{1}{n}\sum_{i=1}^{n} w_i w_i^\tau \xrightarrow[n\to\infty]{} R > 0$，并且

$$\frac{1}{n}\sum_{i=0}^{n}[\|y_i\|^2 + \|u_i\|^2] = O(n^\delta),$$

$$\delta \in \left(0, \frac{1-2\varepsilon(t+1)}{2t+3}\right), \tag{16}$$

那么

$$\|\theta - \theta_n\|^2 = O\left(\frac{\log n(\log\log n)^c}{n^\alpha}\right), \forall\, c > 1,$$

$$\alpha \in \left[\frac{1}{2}(1+\delta), 1 - (t+1)(\varepsilon+\delta)\right].$$

这个定理不仅证明了 $\theta_n \xrightarrow[n\to\infty]{} \theta$，而且还给出了收敛速度. 对随机梯度算法也可以给出类似结果. (16) 是对输入输出讯号增长速度的限制，虽然它允许 $\|y_n\|^2$ 和 $\|u_n\|^2$ 随时间增长，但它们的时间平均值增长不能快于 n^δ，所以进一步的问题是针对具体性能指标所设计的控制规律 u_n^s，是否真的能使诸如(16) 所要求的增长速度得到满足？对二次性能指标和跟踪已知信号这两类指标答案是肯定的. 我们以跟踪为例，设 $\{y_n^*\}$ 是一个已知的有界确定性讯号，要求设计控制，使系统的输出 y_n 尽可能接近 y_n^*，也就是要求

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(y_i-y_i^*)(y_i-y_i^*)^\tau$$

达到极小.

从 $r=1$ 的情形容易看出，$\theta_n^\tau\varphi_n$ 是对 y_{n+1} 的一个估计，所以一个自然的想法是取 u_n^s 使

$$y_{n+1}^*=\theta_n^\tau\varphi_n. \tag{17}$$

我们采用随机梯度算法，取扰动 $\{v_n\}$ 为相互独立的序列，并与 $\{w_n\}$ 独立：

$$Ev_n=0, Ev_nv_n^\tau=\frac{1}{(\log n)^\epsilon}I, \|v_n\|^2\leqslant\sigma^2/\log\epsilon_n,$$

$$\epsilon\in\left(0,\frac{1}{4s(m+2)}\right), s=\max(p,q,r+1), \sigma^2 \text{ 为常数.}$$

实际上系统的输入为

$$u_n=u_n^s+v_n.$$

定理 4 设 1) $\{w_n,F_n\}$ 为鞅差序列，$E(\|w_{n+1}\|^2/\mathscr{F}_n)\leqslant\sigma^2$，$E\sigma^2<\infty$，$\frac{1}{n}\sum_{i=1}^{n}w_iw_i^\tau\underset{n\to\infty}{=}R>0$； 2) $A(z),B(z),C(z)$ 无左公因子，A_p 满秩； 3) $C(e^{i\varphi})+C^\tau(e^{-i\varphi})-I>0, \forall\varphi\in[0,2\pi]$，4) $m=l$，$\det B(z)$ 的零点在单位闭圆外. 那么由(17)确实可以定出 u_n^s，并有

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(\|y_i\|^2+\|u_i\|^2)<\infty,$$

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(y_i-y_i^*)(y_i-y_i^*)^\tau=R \quad (=\min),$$

$$\theta_n\xrightarrow[n\to\infty]{}\theta, a.s.$$

这里我们做到了使性能指标达最优，同时又使参数估计一致，对二次性能指标我们也能同时做到这两点. 但这是两种最简单的指标，对一般的指标目前还未涉及. 但即使是简单的指标，也仍有疑难问题. 就以跟踪而论，定理 4 是以随机梯度算法做的，但在实

际应用中，通常用广义最小二乘算法而不是随机梯度算法，并且也不加扰动 v_n. 当 B_1 已知时，这就是著名的Åström-Wittenmark 自校正跟踪器，它在工程中有许多成功的应用，但它的收敛性、最优性以及精确的收敛速度，只是 1991 年以后的结果，离问题的最初提出已近 20 年了. 并且当 B_1 未知时，问题也没有彻底解决.

对 ARMAX 模型的阶估计，没有像系数估计那样发展得成熟，这是由于反馈控制 $\{u_n\}$ 所造成的困难而导致的. 现在我们还只能给出离线的阶一致估计.

在实际应用中，真实系统往往有别于 ARMAX 过程，或者说 ARMAX 过程只是对真实系统的一种近似，两者之间的差别叫未建模的动态特性. 研究它对系统稳定性的影响，对参数估计的影响，对适应控制效果的影响，不仅在应用上具有十分重要的意义，而且理论上也是一个困难而具有吸引力的课题，这叫稳健性分析. 对未建模的动态特性 η_k 做一些假设，例如要求

$$\|\eta_k\| \leqslant \varepsilon \sum_{i=1}^{k} a^{k-i}(\|y_i\| + \|u_i\| + \|w_i\| + 1), 0 < a < 1,$$

这里 ε 是小常数，但我们没有要求 $\|\eta_k\|$ 本身是一个小量，因为它可以随 $\|y_k\|, \|u_k\|, \|w_k\|$ 的增长而增长. 这时可以证明，广义最小二乘给出的参数估计误差的量级是 ε，在一定条件下，甚至可以做到跟踪误差也是一个小量. 这一类课题要深入研究，要去掉一些复杂的条件，扩大适用范围，特别要结合适应控制的原始出发点，随着系统随时可能发生的变化，即使是缓慢的变化，给出适时的控制律，使系统达到所希望的特性. 对这样一个从实际中提出来的问题，可以说至今没有一个理论上有严格证明的结果.

从上面以 ARMAX 模型为范例对辨识和控制的叙述中，我们已经看到有许多有待解决的困难问题. 但整个控制领域，远比 ARMAX 模型所能概括的要广得多，以随机系统而论，连续时间的适应控制就远没有离散时间那样做得深刻. 参数和阶都已知的

随机系统的控制问题，叫随机控制. 在它的发展过程中最引人注目的是对线性随机系统滤波问题的解决，即 Kalman-Bucy 滤波，以及在二次指标下最优随机控制的求得. 但一旦涉及到非线性随机系统，就远没有那么顺利. 当然有许多重要结果，例如得到了非线性滤波方程，即得到了滤波值应满足的随机微分方程，也得到了条件密度应满足的偏微分方程. 但困难的是，只有一些包括线性、条件正态在内的较简单的情形，这些方程才能解出来，而对一般情形，它们是一组无穷维的方程，实际上解不出来. 用随机过程及微分几何为工具，研究非线性滤波问题，仍是一个困难而吸引人的问题，而非线性随机控制，只有很特殊的情况下，才有完美的解析解. 所以现在对非线性滤波及控制的近似方法及数值方法引起了人们很大兴趣，特别是现代计算机技术及平行计算的发展，为解决非线性滤波问题，提供了新的可能.

在实际应用中出现的系统不仅可能伴有随机干扰，而且往往是非线性的，例如姿态控制，机器人控制及化工过程等. 对确定性的非线性系统，近十多年来，以微分几何为工具作了大量研究. 首先是研究了非线性系统的能控、能观测性，但对工程界带来实质影响的是干扰解耦，输入-输出解耦，反馈线性化这一类问题的解决. 反馈线性化就是用状态反馈的办法，把非线性系统转化为线性系统，这样就可以用一整套成熟的线性控制的工具来解决非线性系统的控制设计问题. 这个方法已成功地用来解决具有多轴线非线性动态的直升飞机的飞行控制问题，并已通过飞行试验. 但对非线性系统的研究，总的来说还处于早期阶段，有许多有待解决的问题，例如要研究系统性能，灵敏度，稳健性，要研制实际可用的包括计算机辅助设计在内的设计方法，要研究输出反馈(至今多用状态反馈)，特别要研究悬而未决的非线性系统的镇定问题，也就是要研究能判别给定的非线性系统可否镇定使之具有一定稳定域的判据. 研究这些问题，不仅要用到深刻的数学工具，并且它们的解决，

对工程实际必定有很大影响.

上面我们已经提到，有些实际系统，例如大型柔性空间天线，加热炉的温度分布，柔性机器人，液流等的运动规律，不仅依赖于时间，也依赖于空间变量，所以它们的动态特性要用偏微分方程或积分-微分方程来描述，这时系统叫分布参数系统. 用泛函分析和偏微分方程为工具，对分布参数系统的研究已取得了许多进展，包括对能观性，能控性，能镇定性的研究，并且已有一些快速的算法来解决分布参数系统的控制问题. 但仍有许多困难问题有待解决，例如：大型柔性结构中的敏感器和执行器本身可对系统的动态特性有很大影响，所以在设计结构和控制系统时，必须把这些因素考虑在内. 抗振是柔性结构中的重要问题，但分布参数系统的内阻尼很不清楚，所以需要研究新的辨识方法以及有效的抗振控制算法；又例如闭环分布参数系统对很小的时滞以及对敏感元件和执行元件位置的很小变动十分敏感，所以灵敏度分析十分重要；当系统中包括液体这一类具有可变边界层时，最优控制和边界摄动的关系，液流控制的有限元近似等都是重要的研究课题. 为了把分布参数控制理论变得更为实用，有许多艰难的数学问题需要解决.

其实，不仅上述随机、非线性、分布参数系统有大量困难的数学问题，即使通常认为解决得较为完整并获得巨大成功应用的多变量线性控制系统，近年来又取得许多进展，H_∞ 优化就是其中一例. 通常系统的性能指标都是从时间域上来考虑的，但实际上，对某些实际问题，联结输入输出的传递函数本身的范数大小可能很重要. H_∞ 优化就是找反馈，不仅使闭环系统稳定，同时要使输入和输出（例如跟踪误差）之间的传递函数的 Hardy 空间的范数最小. 此外，诸如系统性能和不确定因素之间的相互影响，几何描述和代数描述之间的关系，大系统和广义系统等方面，都在不断取得进展. 近年来，多变量控制系统又提出了一些新的重要课题. 系统稳健性要求对象和扰动在较大范围内变化时系统仍保持稳定. 特

别要考察系统的结构稳健性，即系统结构中包括不确定因素时，仍要保证系统设计要求. 确定性系统的适应控制是达到结构稳健性的一种办法，也就是在线地消除结构的不确定性. 但当系统有未建模的高频动态特性及未量测到输出扰动时，某些适应控制算法本身可能变为不稳定的，所以需要研究稳健的适应控制. 希望研究能用较少的未知参数来设计适应控制，如果能有非参数的适应控制，则更理想. 故障容许控制系统是重要的新课题，例如一个系统的部分元件出了故障，飞机的一部分受了损伤，计算机的一部分硬件出了毛病，虽然性能较差，但仍希望保持系统稳定地运转. 设计这样的控制在许多领域有直接的应用，但至今还缺少方法.

以上几种类型系统的模型都用确定性或随机的差分或微分方程来描述，但有时这种模型不适用，例如柔性生产系统，交通系统，计算机通信网络等. 这类系统称为离散事件动态系统. 离散事件就是指某个任务的完成，某个信息的到达等. 现在用有限状态马氏(Марков)链，Petri 网络，排队论及扰动分析法分别对一些特定类型的离散事件系统取得成功，但事实上，至今对离散事件动态系统还缺乏真正好的动态模型和有效的分析方法. 例如，给定一个离散事件系统和它的性能特征，如何设计控制器去满足对系统特性的要求，就是一个十分重要的研究课题，它包括确定满足要求的控制是否存在，当存在时要用有效的算法去求它，还要研究它是否在某种意义下最优的. 这类系统的分析在实际中极为有用，但理论发展还处于初级阶段.

控制理论提出的概念、理论和方法已越出了在工程技术领域的应用，有的在生物、心理及经济等领域中已起到了明显的作用. 辨识和控制的方法在计算机视觉，地震信号处理，物理探矿中的逆问题，多敏感器校正，多目标跟踪，心电图分析，二维图象处理，医学图象处理等方面，正在起着重要的作用.

综上所述，辨识和控制有很强的应用背景，并已取得了许多成

功的应用，但还有大量的有待解决的问题.这些问题的解决，将有力地促进工程技术和其他领域的发展.解决辨识和控制问题，要用到包括微分方程，微分几何，泛函分析，代数逻辑，概率统计，数值分析，复实分析等多种数学工具.当然解决某类特定的辨识和控制问题，就要侧重用到上述数学学科中的某几种，例如解决随机系统问题，较多地用到概率统计和实分析，解决分布参数系统问题，较多地用偏微分方程和泛函分析，而解决非线性问题就较多地用到代数几何及微分方程的方法，等等.总之，研究辨识和控制是应用数学工作者大有可为的场所，许多实际上急待解决而数学上难度很大的问题，正在等待我们去攻克.

参 考 文 献

1. 现代控制理论小丛书，关肇直主编，科学出版社.从1975年起至1988年，已出了十余本，内容涉及现代控制理论的各个方面.
2. P. E. Caines, Linear Stochastic Systems, Wiley, New York, 1987.
3. H. F. Chen, Recursive Estimation and Confrol for Stochastic Systems, Wiley, 1985, New York.
4. L. Ljung and T. Söderstorm, Theory and Practice of Recursive ldentification, MIT Press, 1983.
5. G. C. Goodwin and K. S. Sin, Adaptive Filtering, Prediction and Control, Prentice-Hall, 1983.
6. Feedback Control of Linear and Nonlinear Systems, Ed. by A. V. Balakrishnan and M. Thoma, Springer-Verlag, 1982. (in the Series of Lecture Notes in Control and Information Sciences.)
7. W. M. Wonham, Linear Multivariable Control : A Geometric Approach, Springer Verlag, 1979, New York.
8. Challenges to Control: A Collective View, IEEE Trans. Autom. Control, AC-32, No. 4, 1987, 275—285.

9. W. H. Fleming, Chair, Future Directions in Control Theory, (A Mathematical Perspective), SIAM, Philadelphia, 1988.

10. H. F. Chen and L. Guo, Identification and Stochastic Adaptive Control, Birkhäuser, 1991, Boston.

1

IX

粒子系统与随机分析

Interacting Particle Systems and Stochastic Analysis

严士健
(北京师范大学数学系)
Yan Shijian
(Department of Mathematics,
Beijing Normal University)

粒子系统简介
An Introduction to Interacting Particle Systems

中国数学发展的若干主攻方向
Some Main Aspects on Development of Mathematics in China

引　言

概率论是现代数学中的一个重要学科. 一方面,它有丰富的数学理论,近年来对其他数学学科(如偏微分方程、微分几何、泛函分析、调和分析、半群等)有深入的相互渗透. 另一方面,其他自然科学(如物理、化学、生物、力学等),技术科学(如通讯、信息论、控制论等),管理科学(如运筹学、排队服务理论等)中的很多问题都可以归结为概率模型,应用概率论和随机过程的理论方法加以研究;并且这些问题也向概率论提出了新的重要研究课题. 粒子系统(Interacting Particle Systems)便是导源于统计物理的这样一个新的概率论分支. 由于它处于蓬勃发展的阶段,所包括的范围并不很明确,研究课题也日新月异. 加之作者知识面和水平所限,介绍的内容以及提法定有很多不妥之处,望行家和读者批评指正. 如果这一介绍能引起国内更多读者,特别是有志献身科学的青年的兴趣,作者就达到目的、心满意足了.

概率论与统计物理的联系可以上溯到 19 世纪统计物理学理论建立之初. “统计物理学所研究的对象是大量分子(包括原子、电子、幅射场等)所组成的体系,它的目的是根据分子运动来解释所观察到的物质的宏观热性质. 它的方法是对分子的微观量求统计平均值. ”(王竹溪(1985) 第 2 页.) 因此,一个世纪以来,统计物理学中经常运用概率(在统计物理中常用几率这一术语)的概念和方法,而概率论学者也常常探讨统计物理中的概率论问题或将概率论用于统计物理,出现了一些重要工作,例如 Boltzmann 方程的建立,输运过程理论的建立,Gibbs 的统计系综理论,Хинчин 著的《统计力学的数学基础》(Khinchin(1949)),Колмогоров 在湍流理论方面的工作,等等. 这些对两个学科都有促进作用,但是似乎应该说是不系统的. 到了 20 世纪 60 年代中期,苏联的 Р. Л.

Добрушин 用近代概率论的方法研究了 Ising 模型的相变问题，随后于 60 年代末 Добрушин(1968)，O. E. Lanford 与 D. Ruelle (1969) 独立地提出了无穷粒子系统的 Gibbs 随机场(Gibbs 测度)，开创了粒子系统发展的进程. 利用 Gibbs 随机场讨论相变(超导即为一种相变) 问题，避免了过去使用长程序参数刻划相变时(参看 §1)，需要对与(有限粒子系统上的)Gibbs 态有关的参数取热力学极限(即让系统的粒子数按某种方式趋于无穷)，而这种极限的意义是不很清楚的. 也许更重要的是 Gibbs 随机场这一新的概念一方面是按照粒子的交互作用的规律给出，另一方面它又给出了一个概率框架使经典统计力学的许多模型的大部分平衡性质得以描述. 稍后从系统的动态角度出发，提出了无穷粒子系统的 Марков 过程与动力系统来加以研究. 从广义的角度看，渗流的研究也可以认为是粒子系统的范围. 发展到现在，粒子系统这一概率分支的背景远远超出了平衡态统计物理，Gibbs 随机场在构造量子场中日益显示其作用；有些模型产生于非平衡统计物理、流体动力学、化学、生物以及其他学科. 从学科上说，似乎可以说粒子系统正在发展成描述粒子(分子、原子、……) 系统的运动一种新的数学工具，它更能反映粒子系统运动的本质. 从目前研究状况及文献看，粒子系统的研究内容丰富深刻，研究方法不断发展. 无论从理论上和应用上，都有很多研究需要进行. 对有志于科学事业的青年读者是一个大有用武之地的研究领域，前景是十分吸引人的.

下面我们将逐一向读者介绍，其内容依次分为以下几节：§1 铁磁 Ising 模型的相变. §2Gibbs 随机场. §3 自旋过程. §4 反应扩散模型.

§1 铁磁 Ising 模型的相变

我们先以较经典的方式就铁磁体 Ising 模型介绍相变的有关

问题.

1. 设 Z^d 表示一切 d 维整点组成的集，称为 d 维整点集，即

(1) $$Z^d := \{x = (x_1, \cdots, x_d) : x_i \text{ 为整数}, i = 1, \cdots, d\},^*$$

S 为 Z^d 的一个子集. 设想在每一 $x \in S$ 上有一(铁磁)粒子(因此有时称 S 为位置集)，每个粒子有两种状态，分别记为 $+1$ 与 -1(它们分别代表物理中的正、负自旋)，于是，整个(S 上的)粒子系统的状态(按物理习惯称为组态)可以用 S 到集 $\{-1,1\}$ 的函数(映射)η 表示，用显示法表示 η 就是集合 $\{\eta(u) : u \in S\}$，$\eta(u) = \pm 1$，S 上的全体组态用 $X(S)$ 表示，即

(2) $$X(S) := \{\eta : \eta : S \mapsto \{-1, 1\}\}.$$

当 $S = Z^d$ 时，记 $X(Z^d)$ 为 X.

对于 S 上的粒子系统给出一个**交互作用势** Φ，它是定义在 S 的非空有限子集类 $\mathscr{S}$ 上的实函数，即

(3) $$\Phi : \mathscr{S} \to R(\text{全体实数集}).$$

其物理意义是 $\Phi(A)$，$A \in \mathscr{S}$，表示 A 上全体粒子之间的相互作用的强度. 用下列式子定义的量

(4) $$H_S(\eta) := -\sum_{A \in \mathscr{S}} \Phi(A) \prod_{u \in A} \eta(u)^{**})$$

表示系统的组态为 η 时的总能量(此和式可能发散)，按物理习惯称 $H_S(\eta)$ 为此系统的 Hamilton 量. 此时系统称为 S 上的 Ising 模型.

2. 若上述的交互作用势 Φ 还满足

---

* 在本文中常用$\{\cdots:\cdots\}$表示集合，其中“:”号前表示此集合的元素，“:”之后表示此集合的元具有的特征性质.

** (4)式右边 $\sum_{A \in \mathscr{S}}$ 表示对一切与 $A \in \mathscr{S}$ 有关的量 $\Phi(A) \prod_{u \in A} \eta(u)$ 的求和式，$\prod_{u \in A}$ 表示对一切与 $u \in A$ 有关的量 $\eta(u)$ 的乘积，以后同样以 $\sum$，$\prod$ 分别表示求和式与求积式，求和(积)项写在符号的右边，符号下方表示各求和(或求积)项的标号.

$$(5) \qquad \Phi(A) := \begin{cases} 0, \text{ 当 } |A| \neq 2, \\ \geqslant 0, \text{ 当 } |A| = 2. \end{cases}$$

其中 $|A|$ 表集 A 的元数，则称系统为（无外场）**铁磁 Ising 模型**. 注意此模型中粒子的交互作用只发生在两个粒子之间，而且两个粒子的交互作用的强度非负.

下面就铁磁 Ising 模型给出相变的经典定义. 首先设 S 为 Z^d 的有限子集，$\beta > 0$，交互作用 Φ 满足(5)，易见下列公式(6)规定了 $X(S)$ 上的一个概率：

$$(6) \qquad \begin{cases} \mu_S^{(\beta)}(\{\eta\}) := Z(S,\beta)^{-1}\exp\{-\beta H_S(\eta)\}, \eta \in X(S); \\ Z(S,\beta) := \sum_{\eta \in X(S)} \exp\{-\beta H_S(\eta)\}. \end{cases}$$

称此概率为 S 上具势 Φ 的铁磁 Ising 模型对参数 β 的平衡态（Gibbs 态），简称为S上的Gibbs态. 在物理中 β 是反比于绝对温度 T 的参数. 按照概率论，规定 $X(S)$ 上的实值函数 $f(\eta)$ 对 $\mu_s^{(\beta)}$ 的数学期望为

$$(7) \qquad E_S^{(\beta)} f := \sum_{\eta \in X(S)} f(\eta)\mu_S^{(\beta)}(\{\eta\}).$$

再设 $\Lambda \subset S$，规定铁磁 Ising 模型在 Λ 上的长程序参量为

$$(8) \qquad m_{\Lambda,S}(\beta) := |\Lambda|^{-2}E_S^{(\beta)}[\sum_{u \in \Lambda}\eta(u)]^2,$$

它表示向量 $\{\eta(u): u \in \Lambda\}$ 中的各分量 $\eta(u), u \in \Lambda$ 在 S 的平衡态下的一致程度（即 $\eta(u), u \in \Lambda$，同为 $+1$（或 -1）的比例的大小），一致程度越高，则此参数值越大. $m_{\Lambda,S}(\beta) = 1$ 当且仅当一切 $\eta(u), u \in \Lambda$ 以概率 1 相同，$m_{\Lambda,S}(\beta) = 0$ 当且仅当 $\eta(u), u \in \Lambda$ 以概率 1 正、负各半出现. 再规定 Z^d 上铁磁 Ising 模型的**长程序参量**

$$(9) \qquad m(\beta) := \lim_{\Lambda \nearrow Z^d} \lim_{S \nearrow Z^d} m_{\Lambda,S}(\beta).$$

上式右边中的 Λ, S 永远满足 $\Lambda \subset S$，S 为 Z^d 的有限子集，$\lim\limits_{S \nearrow Z^d}(\lim\limits_{\Lambda \nearrow Z^d})$ 表示 $S(\Lambda)$ 逐步扩大到 Z^d 的极限.

Z^d 的铁磁 Ising 模型相变就是 β 的函数 $m(\beta)$（实际上也是绝

对温度 T 的函数）的解析性质的一种突然改变，确切地说，相变在 $\beta=\beta_c$ 时发生是指：如果 $m(\beta)$ 在某一以 β_c 为端点的开区间上解析，但在包含 β_c 的任何开区间上不等于此解析函数. 特别，如果存在 $\beta_c\in(0,\infty)$ 使

$$(10)\qquad m(\beta)=\begin{cases}0, & \text{当 } \beta<\beta_c,\\ >0, & \text{当 } \beta>\beta_c,\end{cases}$$

则称模型**具有序-无序相变**. 显然有序-无序相变是上述相变概念的一种特殊情形，因为此时 $m(\beta)$ 在 $(0,\beta_c)$ 上等于恒取0值的解析函数，但是在任何包含 β_c 的开区间上不等于0解析函数. 由(6)，(8) 两式易知有限铁磁 Ising 模型不会出现相变，所以相变是无穷粒子系统的一种特有的集体性态.

对于有序-无序相变可以粗略地解释为某种对称性破缺 (broken symmetry). 当 $\beta<\beta_c$ 时，$m(\beta)=0$，由(9) 可见当 Λ 很大时，$\Lambda\subset S$，$m_{\Lambda,S}(\beta)$ 相当小，这就是说 Λ 上的粒子的正、负自旋以很大的概率几乎各半，这可以认为是一种对称性. 当 $\beta>\beta_c$ 时，$m(\beta)>0$. 则表现为：或者正自旋明显多于负自旋，或者反之. 这可以认为是正、负自旋的个数出现了不对称的情况. 这是相变的一种物理理解，在下一节我们还将进一步阐述.

3. 有无相变的问题和在有相变的前提下，**临界值** β_c 的估计或计算问题，都是相变研究中的重要问题. 实际上，超导体的转变温度 T_c 就是某种模型的相变临界值 β_c 的反比函数 $\frac{1}{k\beta_c}$，其中 k 为统计物理中的 Boltzmann 常数，由此也可见研究相变及其临界值的估计的重要性. 作为理论上一个典型的简单模型，人们长期研究了铁磁 Ising 模型的相变问题，下面我们介绍有关的部分著名结果及一些未解决的问题.

设 Z^d 上的交互作用势 Φ 还满足：

$$(11)\qquad \Phi(A):=\begin{cases}1, & \text{当 } A=\{u,v\},\ |u-v|=1 \text{ 时},\\ 0, & \text{当 } A \text{ 为其他子集时},\end{cases}$$

其中 $|u-v|$ 表示 $v,u\in Z^d$ 的距离，即若 $u=(u_1,\cdots,u_d),v=(v_1,\cdots v_d)$，则规定 $|u-v|:=\left[\sum_{i=1}^{d}(u_i-v_i)^2\right]^{1/2}$. 满足(11)的铁磁 Ising 模型称为紧邻 Ising 模型. 已经证明：当 $d=1$ 时，它无相变；当 $d\geqslant 2$ 时，有(有序-无序)相变. 而且当 $d=2$ 时，相变的临界值

$$\beta_c^{(2)}=\frac{1}{2}\log(1+\sqrt{2})\approx 0.44. \tag{12}$$

但当 $d\geqslant 3$ 时，仅仅知道 $\beta_c^{(3)}\leqslant\beta_c^{(2)}$. 求出或较精确估计 $\beta_c^{(3)}$ 的值是一相当困难而有重要意义的研究课题.

对于一维(即 Z^1 上)的铁磁 Ising 模型，若它的交互作用势除满足(5)外还满足平移不变性，即对任何 $u,v\in Z^1,u\neq v$，有

$$\Phi(\{u,v\})=J(|u-v|), \tag{13}$$

其中 $J(n)$ 为定义在正整数集上的函数. 则已经证明：若 $J(n)$ 满足

$$\lim_{N\to\infty}\frac{1}{\sqrt{\log N}}\sum_{n=1}^{N}nJ(n)=0, \tag{14}$$

则模型无相变[32]；若

$$\inf_n n^2J(n)>0, \tag{15}$$

则模型具有序-无序相变(Fröhlich, Spencer(1982)).

有关本节的内容还可参看 Dyson(1971)，那里还提出了一些更深入的问题.

§2 Gibbs 随机场

本节将通过分析长程序参量 $m(\beta)$ 的定义过程来理解并介绍 Gibbs 随机场的概念，给出相变的概率定义，进而介绍 Gibbs 随机场的一些进展情况、研究课题和文献.

1. 在 §1 中定义 $m(\beta)$ 时，首先要用到极限

$$\lim_{S\nearrow Z^d}m_{\Lambda,S}(\beta). \tag{1}$$

而 $m_{\Lambda,S}(\beta)$ 是函数 $[|\Lambda|^{-1}\sum_{u\in\Lambda}\eta(u)]^2$, $\eta\in X(S)$, 关于有限集 S 上的平衡态 $\mu_S^{(\beta)}$ 的数学期望. 由此可以直观地想象(1)式是上述函数对"$X=X(Z^d)$ 上的平衡态"的数学期望. 因此问题转化为如何定义"X 上的平衡态"的问题.

一个最粗糙的想法是: 这种"X 上的平衡态"就是 $\mu_S^{(\beta)}$ 当 $S=Z^d$ 的情况. 但是问题马上发生了. 我们知道 $\mu_S^{(\beta)}$ 是在 §1 中用公式(6)定义的, 在(6)式中当 S 为有限集时是有意义的(归一化常数 $Z(S,\beta)$ 只是 $2^{|S|}$ 项的和式, 因而有意义), 而 $S=Z^d$(无限集)时, $Z(S,\beta)$ 是一个由不可数无穷项组成的和式, 因而不可能定义, 当然 $\mu_Z^{(\beta)}$ 就更不可能用 §1 的公式(6)来定义了. 这就告诉我们"X 上的平衡态"必须用另外的办法定义, 同时也说明(1)的概率意义并不十分清楚!

为了定义"X 上的平衡态", 我们再进一步分析 $m_{\Lambda,S}(\beta)$: 对于任意给定的 $\xi\in X(\Lambda)$, $\zeta\in X(S\backslash\Lambda)$, 用 $\xi\cup\zeta$ 表示 $X(S)$ 的元, 它在 Λ 与 $S\backslash\Lambda$ 上的限制分别是 ξ 与 ζ, 即

$$(\xi\cup\zeta)(u):=\begin{cases}\xi(u), & \text{当 } u\in\Lambda \text{ 时},\\ \zeta(u), & \text{当 } u\in S\backslash\Lambda \text{ 时}.\end{cases}$$

再用 η_Λ 表示函数 $\eta\in X(S)$ 在 Λ 上的限制. 于是用 $X(S)$ 的定义及初等条件概率的定义可以将 $m_{\Lambda,S}(\beta)$ 如下地写出

$$\begin{aligned}m_{\Lambda,S}(\beta)&=\sum_{s\in X(S\backslash\Lambda)}\sum_{\xi\in X(\Lambda)}\Big[|\Lambda|^{-1}\sum_{u\in\Lambda}\xi(u)\Big]^2\mu_S^{(\beta)}(\{\xi\cup S\})\\&=\sum_{s\in(S\backslash\Lambda)}\Big[\sum_{\xi\in X(\Lambda)}(|\Lambda|^{-1}\sum_{u\in\Lambda}\xi(u))^2\mu_S^{(\beta)}(\{\eta:\eta_\Lambda=\xi\}|\\&\qquad\{\eta:\eta_{S\backslash\Lambda}=\zeta\})\Big]\cdot\mu_S^{(\beta)}(\{\eta:\eta_{S\backslash\Lambda}=\zeta\}).\end{aligned}$$

而由 §1 的(6)式及(4)式可以算出(实际下面的计算并不需要 §1(6) 式中的 Φ 满足 §1(5) 式)*: 对任何 $\xi\in X(\Lambda)$, $\zeta\in X(S\backslash\Lambda)$ 有

---

* 导出(3), (4)的计算过程可参看[31]的中译本的译者序.

(3) $$\mu^{(\beta)}(\{\eta:\eta_\Lambda=\xi\}\mid\eta_{S\setminus\Lambda}=\zeta)=\mu^{(\beta)}_{\Lambda,\zeta}(\{\xi\}),$$

其中右边的 $\mu^{(\beta)}_{\Lambda,\zeta}$ 由下列公式给出：

(4) $$\begin{cases}\mu^{(\beta)}_{\Lambda,\zeta}(\{\xi\})=Z^{-1}(\beta,\Lambda,\zeta)\cdot\\ \qquad\qquad \exp\Big\{\beta\sum\limits_{\substack{\Lambda\cap A\neq\varnothing\\ A\in\mathscr{S}}}\Phi(A)\prod\limits_{u\in A}(\xi\cup\zeta)(u)\Big\},\\ Z(\beta,\Lambda,\zeta):=\sum\limits_{\xi\in X(\Lambda)}\exp\Big\{\beta\sum\limits_{\substack{\Lambda\cap A\neq\varnothing\\ A\in\mathscr{S}}}\Phi(A)\prod\limits_{u\in A}(\xi\cup\zeta)(U)\Big\}.\end{cases}$$

(注意：(3) 式左边的 $\mu^{(\beta)}$ 本应为 $\mu^{(\beta)}_S$，右边的 $\mu^{(\beta)}_{\Lambda,\zeta}$ 也应与 $\mathscr{S}$ 有关，但都未明显标出，这是为了后面叙述方便.)

上面的(3) 式只是就 S 为有限集的情形导出的. 但是我们注意：当 $S=Z^d$ 或其他可数集时，若交互作用势 Φ 满足下列条件：对任何 $u\in S$，有

(5) $$\sum_{u\in A\in\mathscr{S}}|\Phi(A)|<\infty.$$

则(4) 式对任何有限集 $\Lambda\subset S$(即 $\Lambda\in\mathscr{S}$) 有意义. 于是(3) 式启发我们如下地给出"X 上的平衡态"的定义.

2. 定义 设 S 为可数集，Φ 为其上的交互作用势且满足(5)，若 $X(S)$ 上的概率测度 μ 使得(3) 对任何有限集 $\Lambda\in\mathscr{S}$ 及任何 $\xi\in X(\Lambda)$，$\zeta\in X(S\setminus\Lambda)$ 成立(其中 $\mu^{(\beta)}_{\Lambda,\zeta}(\{\xi\})$ 由(4) 定义)，则称 μ 为关于 Φ 及 β 的 Gibbs 随机场(Gibbs 测度). 关于给定 Φ，$\beta>0$ 的全体 Gibbs 随机场组成的集记作 $\mathscr{G}(\beta\Phi)$.

上述定义的陈述方式是为了让更多的读者理解 Gibbs 随机场的实质而不牵涉更多的数学概念和术语. 按照数学的精确定义的要求，还需补充以下三点：

(i) $X(S)$ 上的概率测度是指定义在包含一切形如

(6) $$\{\eta\in X(S):\eta(u)=\varepsilon\},u\in S,\varepsilon=+1 \text{ 或 } -1$$

的事件的最小 σ 域(记作 $\mathscr{F}(S)$) 上的概率测度；

(ii) (3) 的左边的精确含意是以测度论为基础的概率理论中的条件概率概念；

(iii) 等式(3)是"μ几乎处处"意义下成立，即除掉一个具μ概率零集以外，等式成立.

可以证明：设Φ是任一满足(5)的交互作用势，β是任一正数，则关于Φ及β的 Gibbs 随机总存在，即 $|\mathscr{G}(\beta\Phi)| \geqslant 1$.

3. 在引进 Gibbs 随机场的概念后，可以给出相变的概率定义，即对满足(5)的给定交互作用势Φ，若 $|\mathscr{G}(\beta\Phi)| = 1$，则称此模型**在$\beta$处无相变**；若 $|\mathscr{G}(\beta\Phi)| > 1$，则称模型**在$\beta$处有相变**. 在一些重要情形(包括铁磁 Ising 模型)下，物理学家采取的相变定义(即§1给出的定义)在多数情形下与此定义一致. 然而最近也发现了两者不一致的例子.

下面介绍Z^d上铁磁 Ising 模型的 Gibbs 随机场的有关结果，以此来说明相变的意义. 设 **1**、**－1** 分别表示Z^d或$Z^d\setminus\Lambda$上恒取值 1，－1 的常值函数. Φ是满足(5)且为Z^d上铁磁 Ising 模型的交互作用势. $\mu_{\Lambda,\mathbf{1}}^{(\beta)}$，$\mu_{\Lambda,-\mathbf{1}}^{(\beta)}$ 是按(4)式定义的相应概率. 可以证明：对任何$\beta > 0$，都存在X上的两个 Gibbs 随机场υ_β^+，υ_β^-，使得对任何$u \in Z^d$，有

(7) $\mu_{\Lambda,\mathbf{1}}^{(\beta)}(\{\eta:\eta(u) = 1\}) \searrow \upsilon_\beta^+(\{\eta \in X:\eta(u) = 1\})$，当 $\Lambda \in \mathscr{S} \nearrow Z^d$；

(8) $\mu_{\Lambda,-\mathbf{1}}^{(\beta)}(\{\eta:\eta(u) = 1\}) \nearrow \upsilon_\beta^-(\{\eta \in X:\eta(u) = 1\})$，当 $\Lambda \in \mathscr{S} \nearrow Z^d$；

(9) $\upsilon_\beta^-(\{\eta \in X:\eta(u) = 1\}) \leqslant \upsilon_\beta^+(\{\eta \in X:\eta(u) = 1\})$；

(10) $\upsilon_\beta^+(\{\eta \in X:\eta(u) = 1\}) + \upsilon_\beta^+(\{\eta \in X:\eta(u) = 1\}) = 1$；

(11) $\upsilon_\beta^+(\{\eta \in X:\eta(u) = 1\})$ 是β的不降函数.

细心的读者会注意到：当u不属于Λ时(7)，(8)两式左边是没有意义的，因为按照(4)式，$\mu_{\Lambda,\mathbf{1}}^{(\beta)}$，$\mu_{\Lambda,-\mathbf{1}}^{(\beta)}$ 是定义在$X(\Lambda)$上的概率. 但是(7)，(8)式中的$\Lambda \nearrow Z^d$，所以当Λ足够大时，就有$u \in \Lambda$，(7)，(8)两式左边就有意义，因此只要我们将(7)，(8)两式理解为对那些包含u的Λ来考虑就可以了. 实际上还有一种办法解释(7)，(8)的

意思，就是将 $\mu_{\Lambda,\zeta}^{(\beta)}$，理解为 X 上的如下测度 L：它在 Λ 内按(4)规定，在 Λ 外集中于点 ζ 上，即对任何 $F \in \mathscr{S}$，定义 F 在 ζ 处的截口 $F(\zeta) \subset X(\Lambda)$ 如下：

$$F(\zeta) := \{\xi \in X(\Lambda) : \xi \cup \zeta \in F\}.$$

再定义

$$\mu_{\Lambda,\zeta}^{(\beta)}(F) := \sum_{\xi \in F(\zeta)} \mu_{\Lambda,\zeta}^{(\beta)}(\{\xi\}), \tag{12}$$

其中 $\mu_{\Lambda,\zeta}^{(\beta)}(\{\xi\})$ 由(4)规定. 于是由(12)定义的 $\mu_{\Lambda,\zeta}^{(\beta)}$ 就是 X 上的概率测度. 可以进一步证明：当 $\Lambda \nearrow Z^d$ 时，$\mu_{\Lambda,1}^{(\beta)}, \mu_{\Lambda,-1}^{(\beta)}$ 分别弱收敛于 $\upsilon_\beta^+, \upsilon_\beta^-$. 还可以证明下述两条结论对铁磁 Ising 模型是成立的：

Ⅰ 当 β 是充分小的正数时，关于 Φ 及 β 的 Gibbs 随机场唯一.

Ⅱ 关于 Φ 及 β 的 Gibbs 随机场唯一的充要条件是 $\upsilon_\beta^+ = \upsilon_\beta^-$，而且这个条件等价于：对任何 $u \in Z^d$，有

$$\upsilon_\beta^+(\{\eta \in X : \eta(u) = 1\}) = \upsilon_\beta^-(\{\eta \in X : \eta(u) = 1\}). \tag{13}$$

关于上述结果的证明读者可参考[40]第三章及附录或[29].

对于铁磁 Ising 模型，令

$$\beta_c := \sup\{\beta > 0 : \upsilon_\beta^+ = \upsilon_\beta^-\}, \tag{14}$$

则由上述结论 Ⅰ 知 $0 < \beta_c \leqslant \infty$. 当 $\beta > \beta_c$ 时，必有 $\upsilon_\beta^+ \neq \upsilon_\beta^-$，所以由结论 Ⅱ 知模型在 β 处有相变. 当 $\beta < \beta_c$ 时，则由 β_c 的定义(14)式知有一 $\beta_1 \in (\beta, \beta_c)$ 使 $\upsilon_{\beta_1}^+ = \upsilon_{\beta_1}^-$，由结论 Ⅱ 的后一部分知对任何 $u \in Z^d$ 有

$$\upsilon_{\beta_1}^-(\{\eta \in X : \eta(u) = 1\}) = \upsilon_{\beta_1}^+(\{\eta \in X : \eta(u) = 1\}).$$

由(10)，(11)知对任何 $u \in Z^d$，$\upsilon_\beta^-(\{\eta \in X : \eta(u) = 1\})$ 都是 β 的不增函数. 于是由 $\beta < \beta_1$ 及(9)式得知对任何 $u \in Z^d$，

$$\upsilon_\beta^-(\{\eta \in X : \eta(u) = 1\}) = \upsilon_\beta^+(\{\eta \in X : \eta(u) = 1\}), \tag{15}$$

再由结论 Ⅱ 的后一部分知 $\upsilon_\beta^- = \upsilon_\beta^+$. 总结以上得到下列的结论

Ⅲ 对任何铁磁 Ising 模型，存在一 $\beta_c \in (0, \infty)$ 使得模型在任何 $\beta < \beta_c$ 处无相变，而在任何 $\beta > \beta_c$ 处有相变(关于 β_c 何时有

限，在 §1 中已有介绍).

关于铁磁 Ising 模型，除了 §1 及上述的结论以外，我们还向读者介绍有关临界指数的一些问题.

为了有关问题的醒目起见，在以下的叙述中，对前面有关参数 β 的结论或符号都通过变换公式 $\beta=\frac{1}{kT}$（其中 k 为一正常数）改用关于参数 T（表示绝对温度）的结论或记号来表达. 例如 $\upsilon_{\beta}^{+}, \upsilon_{\beta}^{-}$ 改记成 $\mu_{T}^{+}=\upsilon_{(kT)^{-1}}^{+}=\upsilon_{\beta}^{+}, \mu_{T}^{-}=\upsilon_{(kT)^{-1}}^{-}=\upsilon_{\beta}^{-}$. 结论就陈述成

Ⅱ′ 对任何铁磁 Ising 模型，存在一 $T_c\in[0,\infty)$ 使得模型在任何 $T>T_c$ 处无相变（即关于 $\frac{1}{kT}\Phi$ 的 Gibbs 随机场唯一），而在任何 $T<T_c$ 处有相变.

下面介绍有关临界指数的问题.

对于 $d=1$，设铁磁 Ising 模型的交互作用势 Φ 满足：

$$\Phi(\{u,v\})=|u-v|^{-\alpha}, u,v \text{ 为整数} \tag{16}$$

（这就是说，§1 的(13)式中的 $J(n)=n^{-\alpha}$, n 为正整数），其中 $0<\alpha<2$. 因此 Φ 满足 §1 的(15)式，从而由该处的结论知存在 $T_c\in(0,\infty)$. 人们猜想下列诸结论成立：存在正常数 β,γ 使得当 $N\to\infty$ 时，

$$E_T^{+}\Big[\sum_{u=1}^{N}\eta(u)\Big]^2\sim\begin{cases}C(T-T_c)^{-\gamma}N, & \text{当 } T>T_c,\\ CN^{\alpha}, & \text{当 } T=T_c;\end{cases} \tag{17}$$

$$D_T^{+}\Big[\sum_{u=1}^{N}\eta(u)\Big]\sim C(T_c-T)^{-\gamma}N, \quad \text{当 } T<T_c; \tag{18}$$

$$E_T^{+}\Big[\sum_{u=1}^{N}\eta(u)\Big]=NE_T^{+}[\eta(0)]\sim C(T_c-T)^{-\beta}, \quad \text{当 } T<T_c, \tag{19}$$

此处 E_T^{+}, D_T^{+} 分别表示关于 μ_T^{+} 的数学期望与方差，β,γ 与 α 无关，称为**模型的临界指数**. 临界指数这种只与一定类型的模型有关而与模型的精细结构无关（例如上述的"与 α 无关"）的特性，在物理

中称为**普适性**，物理中很重视规律的普适性.

对于 $d \geqslant 2$ 的铁磁 Ising 模型的临界指数的存在性，人们也有一系列猜想. 这些都是一些有待解决的既有物理意义又有数学深刻性的重要问题. 人们只是就某些极特殊情况进行了研究，关于这方面的研究情况及问题的进一步叙述可参看[1].

与此有关，在渗流理论中也出现临界指数以及它们之间的标度关系式的存在问题，有兴趣的读者可参看[24].

还应该提及的是：在[1]中还介绍了有关 Gibbs 随机场的中心极限定理的猜想，[46]，[3]的第一章回答了其中一种情形.

4. 现在我们对 §1 所说的对称性破缺作更明确的解释. 当 $T > T_c$ ($\beta < \beta_c$) 时，模型在 T 处无相变，此时只有唯一的平衡态 (Gibbs 随机场) $\mu_T^+ = \mu_T^-$，而且对任何 $u \in Z^d$，由(13) 及(10) 知

$$\mu_T^+\{\eta \in X : \eta(u) = 1\} = \mu_T^-\{\eta \in X : \eta(u) = 1\} = \frac{1}{2}.$$

再由显然成立的等式(因为 $\mu_T^+(X) = 1$!)

(20) $\quad \mu_T^+\{\eta \in X : \eta(u) = 1\} + \mu_T^+\{\eta \in X : \eta(u) = 1\} = 1$

得知

(21) $\quad \mu_T^+\{\eta \in X : \eta(u) = +1\} = \mu_T^+\{\eta \in X : \eta(u) = -1\} = \dfrac{1}{2}.$

当 $T < T_c$ 时，模型在 T 处有相变，此时至少有两个平衡态(Gibbs 随机场) μ_T^+，μ_T^- 即 $\mu_T^+ \neq \mu_T^-$. 由结论 Ⅱ 的后一部分及(9) 式知：存在一个 $u \in Z^d$，使

$$\mu_T^-\{\eta \in X : \eta(u) = 1\} < \mu_T^+\{\eta \in X : \eta(u) = 1\},$$

再由(10) 及(20) 两式即得

(22) $\quad \mu_T^+\{\eta \in X : \eta(u) = -1\} < \dfrac{1}{2} < \mu_T^+\{\eta \in X : \eta(u) = 1\}.$

由于(20) 中的 μ_T^+ 换成 μ_T^- 仍然成立，所以由(10) 即得

(23) $\quad \mu_T^-\{\eta \in X : \eta(u) = -1\} > \dfrac{1}{2} > \mu_T^-\{\eta \in X : \eta(u) = 1\}.$

总结以上讨论得知：当绝对温度 T 较高时(即 $T > T_c$)，模型

只有唯一的平衡态(Gibbs 随机场)$\mu_T^+ = \mu_T^-$,而在每一$u \in Z^d$处粒子的正负自旋出现的可能性各半(即概率都是$\frac{1}{2}$),这是一种对称性.当温度由较高($T > T_c$)变到较低($T < T_c$)时,这个唯一的平衡态出现了分岔,有两个极端的情况μ_T^+,μ_T^-,它们是不同的,而且至少在一$v \in Z^d$处,对于平衡态μ_T^+(相应地对应于μ_T^-)来说,由(22)(相应地(23))知粒子的正自旋的出现的可能性大于(相应地小于)负自旋出现的可能性,而且由(11)知,当T越小,$\mu_T^+\{\eta \in X:\eta(v) = 1\} - \mu_T^+\{\eta \in X:\eta(v) = -1\}$(同样地$\mu_T^-\{\eta \in X:\eta(v) = -1\} - \mu_T^-\{\eta \in X:\eta(v) = 1\}$)越大.这就是说当$T < T_c$时,上述对称性破缺;而且当温度越低,对称性的破缺越大.

5. Gibbs 随机场的概念可以在很一般的情形下引进.其中模型位置集不必是Z^d(这是最本质的情形),而可以是任意可数集.每个位置上的状态空间Y可以是(i)元数大于2的有限集,(ii)紧距离空间,(iii)n维实空间R^n或局部紧距离空间,或者更一般的空间,这里Y是某种代数结构(例如某种群)的情形也许是重要的.还有模型的位置甚至是R^n(或局部紧距离空间)上的局部有限整值测度,这实际上意味着系统的粒子不必位于某种晶格上.对于Gibbs 随机场的概念也不必由交互作用势Φ来定义,而是从更一般地规范(Specification)概念来定义.这种一般的Gibbs 随机场的概念为平衡态统计物理提供了一个理论研究的框架.

由相变的定义可以看出,Gibbs 随机场的存在性、唯一性和非唯一性是它的理论研究的基本问题之一.特别是非唯一性的研究是十分深刻而有重要意义的研究课题.近年来,应用Gibbs 随机场来研究量子场论的某些模型也日益引起数学家和数学物理学家的兴趣,是一个值得注意的研究方向.

6. Gibbs 随机场的研究情况在这篇文章中不可能作进一步的介绍,有兴趣的读者可以参考下列专著并从中找到有关文献.

[31] 是一本着重从一般的规范概念出发，介绍 Gibbs 随机场的概率理论的专著.

[33] 着重介绍相变的数学理论，介绍了一些有物理意义的模型的相变研究，介绍了研究相变的两个重要方法 —— 围道方法及重正化群方法的数学理论.

[19] 系统介绍了近 20 年来发展起来的 Gibbs 随机场与相变的数学理论的基本原理及其应用，列举了相当详尽的文献. 在本书中还介绍了研究相变的另一种重要方法 —— 反射正性的方法. 叙述上也较易阅读，是一本值得推荐的高级入门书.

在[33],[19]中，都引了 Gibbs 随机场在量子场论研究中的应用的一些文献.

[16] 介绍了应用概率大偏差理论研究 Ising 模型的方法及其他有关问题.

§3　自旋过程

1. 问题的提出. 对于一个反映物理现象的数学模型的正确性问题，一方面从直观上判断它的合理性，而且要看从这个模型得出的结果是否具有在物理实验基础上应该具有的性质. 另一方面，就统计物理的平衡态(Gibbs 随机场)来说，当温度给定时，它应该是某种随时间演化的动态系统的不变概率测度(平稳分布). 从这个角度出发，人们不但对 Ising 模型给出了一种无穷粒子 Марков 过程使得 Gibbs 随机场是它的不变测度，而且还给出了更一般的无穷粒子 Марков 过程，它包括了一些非平衡统计物理中的系统. 本节将介绍有关自旋过程的问题，最后简单介绍一下排它过程.

2. 自旋模型. 设 $S=Z^d$ 表示粒子系统的位置集，对于每一 $u\in S$ 有两个状态 0 与 1，于是整个粒子系统的组态空间是

$$X:=\{\eta:\eta:S\to\{0,1\}\}. \tag{1}$$

注意 § 2的组态空间可以如下地归结为这种情形：设$\xi: S \to \{-1,1\}$，即对每一$u \in S$，$\xi(u) = \pm 1$，令$\eta(u) = \frac{1}{2}(\xi(u)+1)$（即$\xi(u) = 2\eta(u) - 1$），则

$$\eta(u) = \begin{cases} 0, & \xi(u) = -1, \\ 1, & \xi(u) = 1, \end{cases}$$

因此$\eta: S \to \{0,1\}$. 而且容易看出：这种变换是一一对应的. 所以§ 2的组态空间可以归结为组态空间(1)来研究. 在本节中X总是指(1)所规定的组态空间.

所谓自旋过程就是一个以X为组态空间随时间t演化的无穷粒子Марков过程，它在$u \in S$处的状态改变速率是$c(u,\eta)$，$u \in S$，$\eta \in X$. 确切地说就是：给定一个非负函数$c(u,\eta)$，$u \in S$，$\eta \in X$. 用η_t表示粒子系统在时刻$t \geqslant 0$的组态，$\{\eta_t: t \geqslant 0\}$是一个随机过程. 再用$P_\eta$，$\eta \in X$，表示在$t = 0$时系统的组态为$\eta$（即$\eta_0 = \eta$），过程$\{\eta_t: t \geqslant 0\}$的分布. 如果$(\{\eta_t: t \geqslant 0\}, \{P_\eta: \eta \in X\})$是一Марков过程而且满足：对任何$u, v \in S$，$u \neq v$，$\eta \in X$，当$t \to 0$时有

$$(2) \quad \begin{cases} P_\eta(\eta_t(u) \neq \eta(u)) = c(u,\eta)t + o(t), \\ P_\eta(\eta_t(u) \neq \eta(u), \eta_t(v) \neq \eta(v)) = o(t). \end{cases}$$

则称$(\{\eta_t: t \geqslant 0\}, \{P_\eta: \eta \in X\})$为以$c(u,\eta)$为速度函数的自旋过程.

自旋模型最为人所知的例子是随机 Ising 模型和基本接触模型，选举模型和近邻模型(Nearest-particle System)有一批学者对它们进行了较完整的研究.

随机 Ising 模型就是 Ising 模型的动态模型（参看[29]第Ⅳ章，[40]第二章 § 3.4). 它的速度函数$c(u,\eta)$满足：对任何$u \in S$，

$$(3) \quad c(u,\eta) \cdot \exp\Big\{\beta \sum_{u \in A \in \mathscr{S}} \Phi(A) \prod_{v \in A} (2\eta(v)-1)\Big\}$$

与$\eta(u)$无关，其中$\mathscr{S}$，Φ的意义与§ 1中的相同. 通常研究速度函数是以下两种：

(4) $c(u,\eta)=\exp\{-\beta\sum_{u\in A\in\mathscr{S}}\Phi(A)\prod_{v\in A}(2\eta(v)-1)\}$

或

(5) $c(u,\eta)=\left[1+\exp\left\{2\beta\sum_{u\in\mathscr{S}}\Phi(A)\prod_{v\in A}(2\eta(v)-1)\right\}\right]^{-1}$.

基本接触模型的速度函数为

(6) $c(u,\eta)=\begin{cases}1, & \eta(u)=1,\\ \lambda\sum_{|v-u|=1}\eta(v), & \eta(u)=0,\end{cases}$

其中 $\lambda>0$ 为常数. 它实际上等价于高能物理中 reggeon 自旋模型，它是 reggeon 场论的一种简化. 它有一些直观有趣的解释. 解释之一是把它解释成传染病模型，当 $\eta(u)=0$ 时，认为在 u 处的个体是“健康”的，而当 $\eta(u)=1$ 时，则认为在 u 处的个体是“病态”的. 于是(6)就可以解释成：当 u 处的个体处于病态时，其治愈率为1；而当 u 处个体健康时，它被传染上疾病的速率与它的带病的紧邻个体数成正比，比例常数为 λ. 解释之二是生死模型：认为 $\eta(u)$ 表示 u 处的粒子数，于是(5)就可解释成：u 处有粒子时，粒子死亡(即 $\eta(u)=1$ 变成 $\eta(u)=0$)的速率是1，v 处的粒子以速率 λ 生出一个粒子并将新生粒子送到紧邻 u(即 $|u-v|=1$)处，若 u 处有粒子则原来粒子与新来粒子合成一个，此外还假定各粒子生出新粒子的行为是概率独立的.

关于选举模型和近邻模型的提法及研究情况可参看[29]第Ⅴ，Ⅵ两章，不再赘述.

3. 对于自旋变相模型，在 $c(u,\eta)$ 满足相当广泛的条件下(参看[29]，[40])，自旋过程是存在的，特别，随机 Ising 过程与基本接触过程是存在的.

基本接触模型是目前已知最简单的具有分岔现象(有时也称非平衡相变)的非平衡系统. 下面我们对它的分岔现象作一直观的介绍.

首先我们注意若 $\eta_t=\theta$，此处及以后 θ 表示各点状态全为0的

组态，则由(6)可以看出对任何$s>0$，必有$\eta_{t+s}=\theta$，即系统的粒子一旦全部消亡，则系统从此以后就不会有任何粒子.再记相应于$\lambda>0$的基本接触过程为$\eta_t^{(\lambda)}$，由λ的含义及(6)式可直观地断言：若$\lambda_1<\lambda_2$，且$\eta_t^{(\lambda_1)}(u)\leqslant\eta_t^{(\lambda_2)}(u)$对任何$u\in S$成立，则$\eta_t^{(\lambda_1)}$在任何$u\in S$处生出粒子都比$\eta_t^{(\lambda_2)}$慢，而死去粒子的速率是一样的.所以人们可以直观地相信下列事实：

(i) 以下用$\eta_t^{(\lambda)}$表示从$\mathbf{1}$出发的而参数值为λ的基本接触过程.若$0\leqslant\lambda_1\leqslant\lambda_2$，且$\eta_t^{(\lambda_2)}$的粒子最终会全部死光，则$\eta_t^{(\lambda_1)}$亦然.因而令

$$\lambda_c=\sup\{\lambda\geqslant 0,\eta_t^{(\lambda)}\text{的粒子最终全部死光}\}\in[0,\infty],$$

则当$\lambda<\lambda_c$时，$\eta_t^{(\lambda)}$的粒子最终会全部死光；当$\lambda>\lambda_c$时，$\eta_t^{(\lambda)}$的粒子永远不会死光.这就是非平衡系统的一种分岔现象.

于是由(i)产生了以下的问题：(a)λ_c是否>0？(b)λ_c是否$<\infty$？(c)能否求出λ_c的精确值或给出λ_c的较精确估计？这些问题是有意义的.下面作些介绍.由(6)知$\eta_t^{(\lambda)}$在任一$u\in S$处由没有粒子变成有粒子的速率

$$\lambda\sum_{|v-u|=1}\eta(v)\leqslant 2d\lambda.$$

而且在多数情况下上式的等号不成立.所以可以直观地相信.

(ii) 若$2d\lambda<1$，则$\eta_t^{(\lambda)}$的粒子最终会全部死光，因而

$$\lambda_c\geqslant\frac{1}{2d}.$$

目前最好的一般结果是

$$\frac{1}{2d-1}\leqslant\lambda_c\leqslant\frac{2}{d}. \tag{7}$$

(以上结果证明参见[29]或[40]，对于$d=1$，还可以略有改进.)

基本接触过程的分岔现象是无穷粒子Марков过程的遍历性的一个问题.设$\mathscr{F}$为包含一切形如$\{\eta\in X:\eta(u)=i\},u\in S,i\in\{0,1\}$的事件的最小$\sigma$-域.$\mathscr{P}(X)$表示$\mathscr{F}$上的一切概率测度组成的集，$(\{\eta_t:t\geqslant 0\},\{P_\eta:\eta\in X\})$为一给定的自旋过程，概率测度

集

$$\mathscr{I}=\left\{\mu\in\mathscr{P}(X):\int P_\eta(\eta_t\in A)\mu(d\eta)=\mu(A),t\geqslant 0,A\in\mathscr{F}\right\}$$

称为此过程的平稳分布(或不变测度)集. 上述的结论(i) 实际上包括在下面的结论中,当$\lambda<\lambda_c$时,$|\mathscr{I}|=1$且对任何$\mu\in\mathscr{P}(X)$都有

$$\int P_\eta(\eta_t\in A)\mu(d\eta)\Rightarrow\delta_\theta,\text{当 } t\to\infty.$$

其中"⇒"表示测度的弱收敛.δ_θ表示集中在单点集$\{\theta\}$上的$(X,\mathscr{F})$上的概率测度,当$\lambda>\lambda_c$时,$|\mathscr{I}|>1$. 对一般的自旋过程,遍历性研究主要是指弄清$\mathscr{I}$的结构、并找出任何平稳分布$\nu\in\mathscr{I}$的吸引域

$$\left\{\mu\in\mathscr{P}(X):\int P_\eta(\eta_t\in A)\mu(d\eta)\Rightarrow\nu,t\to\infty\right\}$$

以及收敛速度等问题. 值得提出的是:由于分岔现象的实际重要性,无穷粒子 Марков 过程的遍历性研究与传统研究的着重点相比有了新的要求. 传统研究着重寻找过程遍历的条件及判别过程遍历的方法,即研究(a) $\mathscr{I}=\{\upsilon\}$为单点集及(b) 对任何$\mu\in\mathscr{P}(X)$都有$\int P(\eta_t\in A)\mu(\mathrm{d}\eta)\Rightarrow\upsilon(t\to\infty)$成立的条件等. 可是对自旋过程及其他无穷粒子系统来说,一个重要的问题是:寻找过程非遍历的条件及判别过程(特别是一些有实际背景的随机过程)非遍历的方法. 从迄今的研究情况看,非遍历的研究有其新特点,一种办法是具体构造一个不同于明显的平稳分布(例如基本接触过程中的δ_θ)的平稳分布或构造两个不同的平稳分布. 也创造了一些方法(如耦合方法、对偶方法、定向渗流方法). 另一办法是估计"最大"与"最小"的平稳分布的耦合距离. 但总的来看,方法不多也不十分有效,这是一个重要的问题,需要人们长期努力.

4. 在第 3 目中提到基本接触模型是一非平衡系统,而随机 Ising 模型是来自平衡态统计物理的一个自旋模型,是一个平衡系

统. 按照物理学家的观点,"平衡等价于可逆". 可以证明随机 Ising 模型是可逆的;即对于给定的交互作用势 $\Phi: \mathscr{S} \to R$ 来说,相应的随机 Ising 模型的速度函数(3) 记作 $c(u,\eta)$. 若 $\mu \in \mathscr{P}(X)$,

$$\int c(u,\eta)[f({}_u\eta) - f(\eta)]\mu(d\eta) = 0, \tag{8}$$

其中 f 为 X 上的任一柱函数,即 f 是只与 η 在某一 $\Lambda \in \mathscr{S}$ 上的值有关的函数,而 ${}_u\eta \in X$ 是将 u 处的值改成 $1-\eta(u)$ 而得到的**组态**,即

$$({}_u\eta)(\nu) := \begin{cases} \eta(\nu), & \nu \neq u, \\ 1-\eta(u), & \nu = u, \end{cases}$$

则称 μ 为 $c(u,\eta)$ 的**可逆测度**. 可以证明:(a)μ 为 $c(u,\eta)$ 的可逆测度的充要条件是 $\mu \in \mathscr{G}(\beta\Phi)$;(b)$c(u,\eta)$ 的任何可逆测度都是 $c(u,\eta)$ 的不变测度;(c) 设 $c(u,\eta)$ 是随机 Ising 模型的速度函数,$\{\eta_t: t \geqslant 0\}$ 是与 $c(u,\eta)$ 相应的随机 Ising 过程,且 η_0 的分布为 $\mu \in \mathscr{P}(X)$(即 η_t 的初分布为 μ),则对任何 $T > 0$,$\{\eta_t: a \leqslant t \leqslant T\}$ 与 $\{\eta_{T-\tau}: 0 \leqslant \tau \leqslant T\}$ 的分布相同的充分与必要条件是 μ 为 $c(u,\eta)$ 的可逆测度. 综合结果(a),(c)就是:μ 为关于 Φ 及 β 的平衡态(Gibbs 随机场)的充分与必要条件是以 μ 为初分布的相应随机 Ising 过程是可逆的. 这就是"平衡等价于可逆"的确切描述. 关于这个问题的完全的解答参看[40]的第二章.

5. 与自旋过程类似,人们还研究了排他过程. 所谓排他过程就是:给定一个非负函数 $c(u,v,\eta)$,$u,v \in S$,$\eta \in X$,而且当 $\eta(u) = \eta(v)$ 时,$c(u,v,\eta) = 0$(此外还满足一些其他条件). 用 η_t,$t \geqslant 0$ 表示粒子系统在时刻 $t \geqslant 0$ 的组态,$\{\eta_t: t \geqslant 0\}$ 是一个随机过程. 再用 P_η,$\eta \in X$,表示在 $t = 0$ 时系统的组态为 η(即 $\eta_0 = \eta$),过程 $\{\eta_t: t \geqslant 0\}$ 的分布. 如果($\{\eta_t: t \geqslant 0\}$,$\{P_\eta: \eta \in X\}$)是一 Марков 过程而且满足条件:对任何三个不同的 $u,v,\omega \in S$,及任何 $\eta \in X$,$\eta(u) \neq \eta(v)$,当 $t \to 0$ 时有

$$(9)\quad \begin{cases} P_\eta(\eta_t(u)=\eta(v),\eta_t(v)=\eta(u))=c(u,v,\eta)t+o(t), \\ P_\eta(\eta_t(u)=\eta(v),\eta_t(v)=\eta(u),\eta_t(w)\neq\eta(w))=o(t), \end{cases}$$

则称$(\{\eta_t:t\geqslant 0\},\{P_\eta:\eta\in X\})$为以$c(u,v,\eta)$为速度函数的**排他过程或粒子运动过程**.

排他过程有一种直观解释如下:设$\eta(u)=0,\eta(u)=1$分别表示在u处没有粒子和有一个粒子两种情况,于是$\eta(u)\neq\eta(v)$分别表示在u,v两处有一处有粒子而另一处没有粒子,不妨设$\eta(u)=1$,因而$\eta(v)=0$,(9)式的第一式就表示系统的组态为η时,u处的粒子移至v处的概率速度为$c(u,v,\eta)$,而第二式表示再有一个位置的粒子数改变的概率速率就是零.对$\eta(u)=0,\eta(v)=1$的情形可作同样说明.按这种解释这个随机过程描述的是一个粒子系统的各个粒子在位置集S的各个位置上移动,一个瞬时只能有一个粒子从它所在位置移到一个空位上,而粒子是不生不灭的.有的文献在适当地选取速度函数$c(u,\nu,\eta)$后,用它作为气体分子运动的简化模型来研究,因而这样一来,就限制气体分子只能在S的元上运动,所以称为**格气模型**.近来这种研究又十分活跃.有可能成为研究现实中气体分子运动的一种有效方法,值得注意.

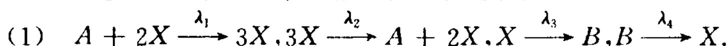
关于自旋过程与排他过程的研究成果及发展状况,可参看[40],[29]及其所引文献,后一书还提出一大批未解决的问题.近期关于格气模型的学术论文,在统计物理杂志(Journal of Statistical Physics)上屡屡出现.我们还将在补注中作些说明.

应用定向渗流方法研究无穷粒子Марков过程的情况可参看[12],[4].近期在杂志上(例如The Annals of Probability)出现了一大批这方面的研究成果.这方面关于基本接触过程的临界现象有下列有趣的结果:设σ_N是在$\{1,2,\cdots,N\}$上的基本接触过程的粒子全部死光的时间,则当$\lambda<\lambda_c$时,σ_N依概率1按$c\log N$增长$(N\to\infty)$;而当$\lambda>\lambda_c$时,σ_N依概率1按e^{cN}增长$(N\to\infty)$.([13],[14].)

§4 反应扩散模型

前面介绍的自旋模型，每个位置上的状态只有两个：0和1，在实际上，有时需要考虑每个位置上有更一般的状态空间. 这一节我们介绍以化学(生物)反应(生灭)为背景的反应扩散模型.

1. 模型的概率假设. 假设我们所要讨论的粒子系统的粒子只一种，用 X 表示，其余参加反应的物种用 $A, B, \cdots$ 表示. 设想它们的反应在一空间区域 V 中进行. 为了简单起见我们先介绍文献上常见的Schlögl模型，在此我们用 X, A, B 表示相应物种的粒子，那么Schlögl模型的反应可以用以下四个式子表达



(1)　$A + 2X \xrightarrow{\lambda_1} 3X, 3X \xrightarrow{\lambda_2} A + 2X, X \xrightarrow{\lambda_3} B, B \xrightarrow{\lambda_4} X.$

(1)的第一式表示两个 X 粒子与一个 A 粒子以概率速率 λ_1 反应成三个 X 粒子，其他意义相同. 除了反应以外，还有粒子的(扩散)运动. 设想将区域 V 分成一些形状和体积相同的小区域(如正方形)，将这些小区域组成的集记成 S. 设所讨论的粒子系统满足以下三条概率假设：

(i) 对于 $u \in S$ 处的几个粒子进行一次反应或 $u \in S$ 处的一个粒子进入 $v \in S$ 处，统称之为系统发生了一次状态改变. 系统在一瞬间发生一次以上的状态改变的概率速率为0；

(ii) $u \in S$ 处的几个特定粒子进行一次第 j 个反应的概率速率是 λ_j；

(iii) $u \in S$ 处的一个粒子进入 $v \in S$ 的概率速率是 $p(u, v)$.

此外还假设每 $u \in S$ 处的 A, B 粒子的个数不变，分别用 a, b 表示.

在这些假设下，用 $\eta(u)$ 表示 $u \in S$ 处的 X 粒子数，则系统的组态 η 就是 S 到 Z_+ (全体非负整数组成的集)的函数，即 $\{\eta(u): u \in S\}, \eta(u) \in Z_+$，其组态空间就是

(2) $$E := \{\eta : S \to Z_+\}$$
$$= \{\eta : \eta = \{\eta(u) : u \in S\}, \eta(u) \in Z_+, u \in S\}.$$
可以证明：在上述假设(i)—(iii)下，Schlögl 模型的组态变化的概率速率由两部分组成：1. 反应部分：在每一 $u \in S$ 处，粒子数由 k 变到 l 的速率是

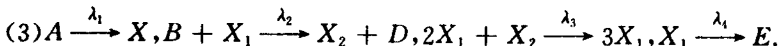
$$q(k,l) = \begin{cases} 0, & |k-l| \geqslant 1; \\ \lambda_1 a C_2^k + \lambda_4 b, & k \geqslant 0, l = k+1, k, l \in Z_+; \\ \lambda_2 C_3^k + \lambda_3 k, & k \geqslant 1, l = k-1. \end{cases}$$

其中 C_i^k 为 k 个不同元素中取 i 个的组合数(当 $i > k$ 时，认为是 0).
2. 扩散部分：设 $u \in S$ 的粒子数为 k，则 $u \in S$ 处有一粒子移至 $v \in S$ 的速率为 $kp(u,v)$.

对于一般模型，也有类似结论. 对于所讨论的粒子系统只有一种粒子的，我们称为单物种的，Schlögl 模型便是单物种粒子系统的一个特殊情形，还可以讨论更一般的情形. 详情见[38].

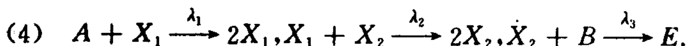
我们称有一种以上粒子的粒子系统为多物种的. 多物种粒子系统最简单而有兴趣的模型是三分子模型(亦称 Brusselator)及弱肉强食模型(亦称 Volterra-Lotka 模型)，它们的反应式如下给出：

三分子模型：



(3) $$A \xrightarrow{\lambda_1} X, B + X_1 \xrightarrow{\lambda_2} X_2 + D, 2X_1 + X_2 \xrightarrow{\lambda_3} 3X_1, X_1 \xrightarrow{\lambda_4} E.$$

弱肉强食模型：



(4) $$A + X_1 \xrightarrow{\lambda_1} 2X_1, X_1 + X_2 \xrightarrow{\lambda_2} 2X_2, X_2 + B \xrightarrow{\lambda_3} E.$$

对于多物种模型也可以求出系统组态变化的概率速率. 详情可参看[38]. 在[38]中还讨论了非线性 Master 方程的模型.

2. 当 S 为有限集时，反应扩散模型就是 $|S|$(表示 S 的元素)维 Q-过程，严士健与陈木法在[39]中，得到了这种一般模型的多维 Марков 链(Q-过程)的存在唯一性、常返性及正常返性的结果.

文中建议了一种将多维 Q- 过程归结为一维 Q 过程来研究的方法. 应用此文中关于正常返性的一般性结果可以证明 Schlögl 模型是遍历的. 但是该结果对三分子模型无效. 韩东在[22]中证明了当 $|S|=1$ 时,三分子模型是遍历的,陈金文在[4]的 § 3.2 中证明了当 S 为任意给定的有限集时,三分子模型也是遍历的.

由于物理学家相当普遍地认为 Schlögl 模型和三分子模型有分岔现象. 从概率论的观点来理解,分岔现象应该理解为它的不变测度不唯一,非遍历. 所以自然可以认为用有限粒子系统来刻划非平衡系统的分岔现象未必恰当. 类比于平衡态统计物理相变的研究,我们认为研究无穷粒子反应扩散模型是适当的.

3. 无穷粒子反应扩散模型的存在性与唯一性. 现在设 S 为可数集,粒子系统的组态空间仍由(1)给出(不过 S 为可数). 设系统组态变化的概率速率仍为反应部分和扩散部分:其中反应部分是在每一 $u \in S$ 处,粒子数由 k 变成 l 的速率是 $q_u(k,l)$. 扩散部分比第1目稍广一点,改为设 $u \in S$ 的粒子数为 k, $u \in S$ 处有一粒子移至 $v \in S$ 处的速率为 $c_u(k)p(u,v)$. 这就是单物种无穷维反应扩散模型,在 $c_u(k)$ 以 k 的线性函数为界的条件下,郑小谷,丁万鼎在[41]中首先应用耦合方法证明反应项为线性增长的情形的存在定理. 陈木法在结合 q- 过程理论与耦合方法证明了更一般的无穷维反应扩散过程的存在,进一步他还结合 KRW- 距离理论证明了更一般的存在定理. 这些结果总结在[6]的第三、六章中,它包括了[29] Ⅸ 章中所列各模型的存在性作为特例.

对于包括三分子模型的多物种无穷维反应扩散模型,唐守正在[34]中建立了 Polish 空间上的极大值原理,用它证明了:在扩散系数有界的条件下,相应的反应扩散过程存在.

对于多物种的一般情形,用的是鞅方法,构造以 $\{\eta:\eta:S \to \{0,1\}\}$ 为组态空间的无穷粒子 Марков 过程的鞅方法可从[29]1. § 6 及其所引文献了解. 李俊平在[26]中,韩东在[21]中,对组态空间

赋予不同拓扑的情况下，证明了在扩散系数以线性函数为界的条件(还有一些关于反应部分的条件)下，反应扩散模型的鞅解是存在的，对于鞅解的唯一性，韩东也得到相当一般的结果. 因而对多物种情形，在相当广泛的条件下，无穷维反应扩散过程是存在的. 与此相关，冯水在[17]，[56]中，用鞅方法证明了单物种非线性 Master 方程的鞅解的存在性和唯一性，因而它是 Mckean 意义下的非时齐 Марков 过程.

上面介绍的是关于无穷维反应扩散过程的存在问题的进展情况，至于无穷维反应扩散过程的唯一性问题，陈木法在[8]中，李勇在[28]，[65]中对单物种情形得到了满意的结果，对多物种情形也得到一些较好的结果. 李勇在讨论唯一性问题的过程中，还改进了 Polish 空间上的极大值原理.

4. 无穷粒子反应扩散模型的遍历性. 陈木法在[6]中及黄力平在[23]中得出一般无穷粒子 Марков 过程的平稳分布存在的充分条件和平稳分布唯一及遍历的充分条件. 丁万鼎，郑小谷在[10]中讨论了单物种具线性增长和死亡速率的反应扩散过程的遍历性问题，得到发生非平衡相变的临界区域；找出了一切平移不变的平稳分布及相应的吸收域. Ding、Durrett，Liggett 在[11]中研究了，$S=Z^d$ 时，多项式反应模型(包括 Schlögl 模型)的可逆情形. 证明了此时平稳分布唯一而且遍历. Neuhauser 在[30]中讨论了上述模型的一般情形(不一定可逆)，证明了当迁移速率 $c_u(k)/k$ 充分小时，过程遍历. 陈木法在[7]中对上述模型的一般情形应用耦合方法得到使过程遍历的很一般的充分条件，包括了 Neuhauser 的结果. 应该指出：上述三个结果中都包含“纯生系数大于零”的条件.

李勇与郑小谷在[27]中研究了无穷维反应扩散过程的非遍历性问题，他们将定向渗流方法推广而称之为有色随机图，并将它应用于以 $\{\eta:\eta:S\to\{0,1,2\}\}$ 为组态空间的无穷粒子系统，得到该系统非遍历的一个充分条件. 然后应用耦合方法得到一类无穷维

反应扩散过程非遍历的充分条件. 作为例子,得到纯生系数等于零的 Schlögl 模型非遍历的充分条件.

由上面的介绍可以看出:对于无穷维反应扩散过程,即令是 Schlögl 模型. 要想得到较完满的遍历性结果,也还需要很多努力. 一个有兴趣的问题是:当纯生系数为充分小的正数时,过程是否非遍历?至于多物种情形的遍历性研究还是空白.

与此相关,冯水在[17],[56]中讨论了非线性 Master 方程的解的分岔现象,证明了它的平稳分布的个数等于一幂级数的零点个数,由此找出了 Schlögl 模型的参数的两个区域,它们分别是使 Schlögl 模型的平稳分布唯一和不唯一的参数区域.

5. 无穷维反应扩散模型的连续性极限(文献上也常称为流体动力学极限 (Hydrodynamical limit), 有的也称为动力学极限 (Kinetic limit)). 最早研究 Schlögl 模型是假设在容器内,所讨论的物种的密度是均匀的. 于是按照化学动力学的反应律列出密度所满足的常微分方程(多物种情形则为方程组),由此常微分方程研究模型的性质,后来考虑到容器内密度的不均匀性,这就发展为应用偏微分方程(反应扩散方程)来研究模型的非平衡统计物理问题. 这一节我们介绍了用无穷维反应扩散模型来研究这一类非平衡相变的问题. 那么很自然也会提出这样的问题:作为同一类物理问题,提出的两种数学模型是否一致呢?这个问题从七十年代末就一直有人关心和探讨. 近些年来,Boldrighini 等在[2]中研究了单物种无穷维反应扩散过程的连续型极限问题. 他们考虑 $S = Z$ 的情形. 当过程的初始分布满足一定条件时,过程的扩散速率和位置经适当地标度(Scaling)后,过程的数学期望的极限(当标度参数趋于零)满足相应的反应扩散方程. 陈木法等在 1989 年[70]研究了某种多物种的情形,徐先进在[37]中讨论了多物种且 $S = Z^d$ 的情形的连续型极限及有关问题.

应该指出的是:无穷粒子系统的连续型极限不仅限于反应扩

散模型，对于自旋模型、排他模型等的流体动力学极限，在 80 年代期间，有相当广泛的研究. 这里就不再赘述了.

补　　注

趁校对之便，补充一些参考文献. 和以前一样，并不为求全，只是作者所了解并认为值得向读者推荐的一些材料，以及原来没有公开发表而最近以论文形式发表的结果. 结合文献目录说明近来的粒子系统及有关论题方面值得注意的一些情况.

1. 我想首先值得向年轻读者推荐的是专著[47]，它是[6]的增补与改写，但是与[6]有很大的改变，可以说是重写. 添加了跳过程与粒子系统方面很多新进展，特别是关于耦合方法的论述相当精彩. 充分显示了这一方法在无穷粒子系统及随机过程研究方面的作用.

2. 格点分形上 Ising 模型的相变. 最近[69]讨论了二维以上(含二维)格点 Sierpinski 海绵和 Sierpinski 地毯上的紧邻 Ising 模型的相变. 作者证明了前者没有相变，而后者有相变. 这两个结果在[58],[59]中用重正化群方法讨论过. [69]用完全不同的 Peierl 围道方法严格地加以证明，在证明过程中，作者对 Sierpinski 地毯的高维情形用一种新的办法构造了对偶图，并引用代数拓扑的一些重要定理(如 Alexander 对偶定理等)来讲清楚道理. 这个结果的二维情形在[47]中有论述.

[67]是最近出版的一本用 Cluster 展开的方法研究 Gibbs 随机场的新书.

3. §3.3 谈到对于基本接触过程，存在数值 λ_c，使得当 $\lambda < \lambda_c$ 时，系统的粒子最终会全部死光；当 $\lambda > \lambda_c$ 时，系统的粒子永远不会死光. 于是出现一个问题：当 $\lambda = \lambda_c$ 时，系统最终会出现什么情况？最近，[44]给出并证明了下列明确的结论：当 $\lambda = \lambda_c$ 时，系统的

粒子最终会全部死光. 在证明中,用到了渗流的重正化方法,应用这个方法,作者们在[44]及其他论文中,还讨论了其他一些问题.

顺便提一句,[61]是最近在国内发行的一本好的有关渗流的专著.

4. §3.5的倒数第二、三两段讲到格气模型. 吉林大学的胡守信教授送给作者文献[57],使作者进一步了解到力学家们关于格气的一些重要的理论工作. [57]是文章作者先前提出的一些模型的理论探讨,模型是一种特殊的每一因子空间有限的时间离散的可数维马链,他们讨论了这种马链的流体动力学极限与流体动力学方程的关系. [54]中也有大量的篇幅讨论格气问题. 近年来,在统计物理杂志(Journal of statistical of physics)中,经常有讨论格气的论文. 甚至有时出专集(很多不是用无穷粒子系统的方法). 总之,格气的研究是近年来一个活跃的研究方向.

[54]是目前比较系统地综述了流体动力学极限的数学研究的一本著作,那里列出了大量的文献,[55]用定向渗流的方法讨论流体动力学极限问题,这些都值得有兴趣的读者参考.

5. 关于无穷维反应扩散过程的遍历性,[48]对Schlögl模型的一个具体情形进一步讨论了它的遍历区域,[66]讨论了一类平移不变的无穷维反应扩散过程的遍历性.

6. 近年来,人们开始研究随机环境下的无穷粒子系统. 例如,[45]中讨论了随机环境下可逆紧邻粒子系统,[60]中讨论了随机环境下的分支随机游动,在Journal of Statistical Physics和The Annals of Probability以及其他杂志上还有一些这类的论文.

与此有关的是研究在随机陷井场中迁移的问题. [63]给出了这个问题的一个综述,[62],[53]对一些情形进行了研究.

7. 最后,我想指出一点:KRW-距离和耦合方法在研究Gibbs随机场和无穷粒子系统的过程中,都是一个有效的方法. 看来,这个方法在数学的更广泛的领域中有用. 最近,[68],[51],[49]第

一次用它估计了流形的第一本征值，这是一个在分析数学中著名的重要数学问题，有相当的难度. 上述论文对于零曲率的情形得到了丘成桐等的最好结果，而在曲率不等于零的情形，则全面地改进了已有的结果. 而且证明要直截了当得多. 由此可见耦合方法的有效性. 此外，他们还改进了流形的梯度估计. 这方面的大部份文章已经或将在中国科学，数学学报上发表.

参 考 文 献

以下所列的参考文献十分不全，主要是文中用到的材料，或者是作者认为有兴趣的读者可以进一步查阅的材料.

1. Bleher, P. M., Major, P. (1987), Critical phenomena and universal exponents in statistical physics on Dyson's hierarchical model, Ann. Prob. Vol. 15, 431—477.
2. Boldrighini, G., DeMasi, A., Pellegrinotti, A., Presutti, E. (1987) Collectiue phenomena in interacting particle systems, Stoch. Proc. Appl. 25, 137—152.
8. 陈冬青(1990)，Gibbs 随机场的中心极限定理，北京师范大学博士学位论文.
4. 陈金文(1991)，粒子系统的若干极限定理，北京师范大学博士学位论文.
5. Chen Jin-Wen(陈金文), Durrett, R., Liu Xiu-Fang(刘秀芳)(1990), Exponential convergence for one dimentional contact processes, Acta Math. Sinica (New Series), 6, 349—353.
6. 陈木法(1986)，跳过程与粒子系统，北京师范大学出版社.
7. Chen Mu-Fa(1990), Ergodic theorems for reaction-diffusion processes, J. Statist. Phys. 58. 939—966.
8. Chen Mu-Fa (1991), Unigueness of reaction-diffusion processes, Chin. Sci. Bulletin, 36. 969—973.
9. Chen Mu-Fa, Huang Li-Ping, Xu Xian-Jin (1991), Continuum limit

for reation-diffusion processes with several species. to appear in Nankai Series if Pure and Applied Mathematics and Theoretical Physics, World Scientific, Singapore(见[70]).

10. Ding Wan-Ding(丁万鼎), Zheng Xiao-Gu(郑小谷)(1989), Ergodic theorems for linear growth processes with diffusion, Chinese Ann. Math. 10B. 386—402.

11. Ding Wan-Ding, Durrett, R., Liggett, T. M. (1990), Ergodicity of reversible reaction-diffusion processes, Probab. Th. Rel. Fields, 85, 13—26.

12. Durrett, R. (1988), Lecture Note in Particle Systems and Percolations, Wadsworth Inc.

13. Durrett, R., Lu Xiu-Fang(刘秀芳) (1988), The contact process on a finite set, Ann. Prob. 16, 1158—1173.

14. Durrett, R., Schonmann, R.H. (1988), The contact process on a finite set. Ann. Prob. 16, 1570—1583.

15. Dyson, F.J. (1972), Existence and nature of phase transitions in one dimensional Ising ferromagnets, in Mathematical aspects of Statistical mechanics, SIAM-AMS Proceedings, Vol. 5. U.S.A.

16. Ellis, R.S. (1985), Entropy, Large Deviations and Statistical Mechanics, Springer-Verlag, Germany.

17. 冯水(1987),一类非线性 Master 方程的解,北京师范大学硕士学位论文.(Feng Shui, Zheng Xiao-Gu(1988), Solutions of a class of nonlinear master equations. Technical Report 115, Carleton University, Ottawa.)

18. Fröhlich, J., Spencer, T. (1982), The phase transition in the one-dimensional Ising model with $1/r^2$ interacticn energe, Comm. Math. Phys., 84, 87—101.

19. Georgii, H.-O. (1988), Gibbs Measures and Phase Transitions, Walter de Gruyter, Germany.

20. 韩东(1989),多物种无穷维反应扩散粒子系统鞅解的存在唯一性,北京师范大学博士学位论文.

21. 韩东(1990),多物种无穷维反应扩散粒子系统鞅解的存在性(英文),应用概率统计,6,265—278.
22. 韩东(1991),非平衡系统中两个概率模型的遍历性,新疆大学学报.
23. 黄力平(1987),一类无穷质点马尔可夫过程的平稳分布存在性,应用概率统计,3, 152-158.
24. Kesten, H. (1987), Percolation theory and first-passage percolation, Ann. Prob. 15, 1231—1271.
25. Khinchin, A. I. (1949), Mathematical Foundations of Statistical Mechanics (Translated from the Russian), Dover Publications, Inc. New York.
26. 李俊平(1988),多物种反应扩散过程,北京师范大学硕士论文.
27. 李勇(1988),关于无穷维反应扩散过程及排它过程的若干结论,北京师范大学硕士论文.
28. 李勇(1991),反应扩散过程的唯一性、随机经济系统的稳定性,北京师范大学博士学位论文.
29. Liggett, T. M. (1985), Interacting Particle Systems, Springer-Verlag.
30. Neuhauser, C. (1990), An ergodic theorem for schlögl models with small migration, Probab. Th. Rel. Fields, 85, 27—32.
31. Preston, C. (1976), Random Fields, LNM 534, Springer-Verlag. (严士健、陈木法、丁万鼎译,随机场,北京师范大学出版社,1982.)
32. Rogers, J. B., Thompson, C. J. (1981), Absence of long-range order in one-dimensional spin systems, J. Statist. Phys., 25, 669—678.
33. Sinai, Ya. G. (1982), Theory of Phase Transitions: Rigorous Results, Pergamon Press.
34. 唐守正(1985),离散无穷粒子系统中的某些问题,北京师范大学博士学位论文.
35. 唐守正(1985),扩散速度有限的多物种反应扩散过程的存在性及唯一性,应用概率统计,1,11—22.
36. 王竹溪(1985),统计物理导论,人民教育出版社,第二版.
37. 徐先进(1991),连续型极限,北京师范大学博士学位论文.

38. 严士健、李占炳(1980),非平衡系统的概率模型及 Master 方程的建立,物理学报,29,139—152.

39. Yan Shi-Jian, Chen Mu-Fa (1986), Multidimensional Q-processes, Chinese Ann. Math. 7B, 90—110.

40. 严士健(1989),无穷粒子马尔可夫过程引论,北京师范大学出版社。

41. Zheng Xiao-Gu, Ding Wan-Ding(1987), Existence theorems for linear growth processes with diffusions, Acta Mathematica Scientia 7, 25—42.

42. Zheng Xiao-Gu, Zhen Wen-Qu(曾文曲)(1987), An ergodic theorem for generalized simple exclusion processes with reversible positive transition, Acta Mathematica Scientia 7, 169—175.

43. Zheng Xiao-Gu(1988), Ergodic theorem for generalized long-range exclusion processes with positive recurrent transition probabilities, Acta Math. Sinica (New Series), 4,193—209.

44. Bezuidenhout, C and Grimmett, G. (1990), The critical contact process dies out, Ann. Prob. 18, 1462—1482.

45. Chen Da yue (陈大岳) and Liggett, T. M. (1992), Finite reversible nearest particle systems in inhomogeneous and random environment, Ann. Prob. 20, 152—173.

46. Chen Dong Ching (陈冬青) (1991), A uniform central limit theorem for nonuniform ϕ-mixing random fields, Ann. Prob. 19, 636—649.

47. Chen Mu Fa (陈木法) (1992), From Markov Chains to Non-equilibrium Systems, World Sci., Singapore.

48. Chen Mu fa (陈木法) (1993), On ergodic region of Schögl's model, CARR Report in Math. Phys. n. 16, 1—15.

49. Chen Mu Fa (陈木法) (1993), Optimal couplings and applications to Riemannian geometry, Probability Theory and Mathematical Statistics, Vol. 1,1994 VPS/TEV (B. Grigelionis et al., eds.).

50. Chen Mu Fa (陈木法) and Wang Feng-Yu (王凤雨) (1993), On order-preservation and positive correlation for multidimentional processes, Prob. Th. Rel. Fields 95, 421—428.

51. Chen Mu Fa (陈木法) and Wang Feng-Yu (王凤雨) (1993), Application of coupling method to the first eigenvalue on manifold, Tech. Rep. Labor. in. Prob. & Statis., No. 215, Carleton Univ., Science in China(A), 37(1994),1—14.

52. Chen Mu Fa (陈木法), Yan Shi-Jian (严士健) (1991), Jump processes and particle systems, Probability Theory and Its Applications in China, AMS Contemporary Math. Series Vol. 118 (Yan, S. J., Yang, C.C. Wang, J.G., eds), pp. 223—57.

53. Chen Mu-Fa (陈木法) Yan Shi-Jian (严士健), Zhou Xian-Yin (周先银) (1992), The range of random walk on trees and related trapping problem, 已投应用数学学报.

54. DeMasi, A., Presutti, E. (1992), Mathematical Methods for Hydrodynamic Limits, Lecture Notes in Math. 1501, Springer-Verlag, Berlin, Heideberg, New York.

55. Durrett, R. and Neuhauser, C. (1991), Paticle systems and reaction-diffusion equations, Tech. Report 92, MSI, Cornell Univ., Ithaca, New York.

56. Feng Shui (冯水) and Zheng Xiao-Gu (郑小谷) (1992), Solutions of a class of nonlinear master equations, Stoch. Proc. Appli. 43, 65—84.

57. Frisch, U., d'Humieres, D., Hasslacher, B., Lallemand, P., Pomeau, Y., (1987), Lattice gas hydrodynamics in two and three dimensions, Complex Systems 1, 649—707.

58. Gefen, Y., Aharony, A., Shapir, Y. and Mandelbrot, B.B. (1984), Phase transitions on fractals: I, Sierpinski gasket, J. Phys. A: Math. Gen. 17, 435—444.

59. Gefen, Y., Aharony, A., Mandelbrot, B.B. (1984), Phase transitions on fractals: II, Infinitely ramified lattices, J. Phys. A: Math. Gen., 17, 1277—1289.

60. Greven, A. and Hollander, F.d. (1992), Branching random walks in random environment: phase transitions for local and global growth rates, Prob. Th. Rel. Fields, 91, 195—249.

61. Grimmett, G. (1989), Percolation, Springer-Verlag, 世界图书出版公司, Berlin, 北京.

62. Hollander, F. d., Shuler, K. E. (1992), Random walks in random field of decaying traps, J. Stat. Phys., 67, 13—31.

63. Hollander, F. d. and Weiss, G. H. (1992), Aspects of trapping in transport processes, Some Problems in Statistical Physics, SIAM.

64. 李勇 (1990), 关于有势排它过程的正可逆测度,应用概率统计, 6, 121—126.

65. 李勇 (1991), 无穷维反应扩散过程的唯一性,科学通报,36, 1681—1684.

66. 李勇 (1993),一类平移不变无穷粒子反应扩散过程的遍历性,预印本.

67. Malyshev, V. A. and Minlos, R. A. (1991), Gibbs Random Fields: Cluster Expansions (Translated from the Russian by Kotecky, R. and Holisky, P.) Kluwer Academic Publishers, Dordrecht, Boston, London.

68. 王凤雨 (1992),流形上的扩散过程及其应用,北京师范大学博士学位论文.

69. 郑君礼 (1993),格点分形上的 Ising 模型的相变, Q- 过程的鞅方法, 北京师范大学博士学位论文.

70. Chen Mu Fa (陈木法), Huang Li Ping (黄力平) and Xu Xian Jin (徐先进)(1992), Continuum limit for reaction diffusion processes with sereral species, Probability and Statistics, Proc. of The Special Programm at Nankai Instifute of Mathematics, Tianjin, China, Aug., 1988—May, 1989, World Sci., Singapore.

严加安
（中国科学院应用数学研究所）
Yan Jiaan
（Institute of Applied Mathematics,
Academia Sinica）

随机分析学简介
A Short Presentation of Stochastic Calculus

中国数学发展的若干主攻方向
Some Main Aspects on Development of Mathematics in China

§1 引　言

诞生于 50 年代的随机分析学是概率论的一个重要分支. 它的创始人伊藤清(K. Itô) 获 1987 年度 Wolf 数学奖. 在对获奖工作评价中写道:"他的随机分析可以看作随机王国中的牛顿定律,它提供了支配自然现象的偏微分方程和隐藏着的概率机制之间的直接翻译过程,其主要成分是对布朗运动函数的微分和积分运算,由此产生的理论是近代纯粹与应用概率论的基石."40 年来,随机分析随着随机过程一般理论及现代鞅论的产生和发展而形成为概率论的一个最富于生命力的分支. 随机分析不仅为概率论及随机过程的理论研究提供了强有力的工具,而且对数学的许多分支(如偏微分方程,调和分析,微分几何)、滤波与控制、通讯与动态系统及金融经济学等有广泛应用. 近十多年来,数学物理(如统计力学、量子力学和量子场论) 对随机分析提出了许多新问题,刺激了随机分析的发展. 反过来,随机分析的发展又为数学物理提供了新的工具和方法. 这两者之间的愈来愈强的交互作用决定了当前随机分析发展的主流方向. 这些方向是: **Malliavin 分析**、**Dirichlet 型**、**白噪声分析**、**大偏差理论**、**无穷维随机分析**及**流形上的随机分析**等. 本文首先介绍**伊藤随机分析**的理论及应用, 然后概略介绍随机分析的若干研究方向,以期引起概率界同行对随机分析学的兴趣和关注.

§2 布朗运动、马氏过程和鞅

随机分析学的研究对象是随机过程. 所谓随机过程,它是描述随时间变化的偶然现象的数学模型. 在随机分析学中,最重要的随机过程是 Brown 运动、Марков 过程和鞅. 其中 Brown 运动尤为重

要. 所谓 Brown 运动，它是对悬浮于液体中胶质粒子作极不规则运动这类物理现象的统称. 这一现象最早由英国植物学家 R. Brown 于 1827 年观测到并提出来研究. 1905 年 Einstein 从概率论及统计力学出发导出了对 Brown 运动的如下统计描述：从 x 出发的 Brown 运动粒子在时刻 t 在 y 处被发现的概率密度 $p_t(x,y)$ 满足如下的扩散方程(热传导方程)：

$$\left(\frac{\partial}{\partial t} - D\sum_{i=1}^{3}\frac{\partial^2}{\partial y_i^2}\right)p_t(x,y) = 0, \qquad \lim_{t\downarrow 0} p_t(x,y) = \delta(y-x). \tag{2.1}$$

即有 $p_t(x,y) = (4\pi Dt)^{-\frac{3}{2}}\exp(-|y-x|^2/4Dt)$. 其中

$$D = \frac{kT}{m\beta},\ \beta = 6\pi\eta a, \tag{2.2}$$

这里 k 为 Boltzmann 常数，T 为液体的绝对温度，m 为粒子质量，a 为粒子直径，η 为液体的粘性系数. 这一统计描述的正确性于 1908 年被 Perrin 精心设计的实验所证实. 但直到 1923 年 R. Wiener 才用随机系数的三角级数构造出满足上述性质的连续随机过程. Wiener 构造出的 d- 维随机过程 (B_t) 具有如下性质：(1) 过程的轨道连续；(2) 对 $t_0 < t_1 < \cdots < t_n$，增量 $B_{t_1} - B_{t_0}, \cdots, B_{t_n} - B_{t_{n-1}}$ 相互独立；(3) $B_t - B_s$ 服从参数为 $|t-s|$ 的正态分布. 如果 $B_0 = x$，则称 (B_t) 为 **从 x 出发的(标准)Brown 运动或 Wiener 过程**. 可以证明 Wiener 过程的几乎所有轨道是处处不可微的.

1906 年 Марков 首次提出了一类非独立的随机变量序列模型，其特点是：已知现在的状态，未来与过去独立. 后来就把具有这种"无后效性"的随机过程称为 Марков **过程**. 设 E 为过程的状态空间，$\mathscr{E}$ 为 E 上的一 σ- 域. Марков 过程的演化规律由 $(E, \mathscr{E})$ 上的一族概率测度 $\{P(s,x,t,\cdot), s < t, x \in E\}$ 决定，其中 $P(s,x,t,\cdot)$ 满足如下的方程：

$$P(s,x,t,A)=\int_E P(s,x,u,\mathrm{d}y)P(u,y,t,A),\ s<u<t. \tag{2.3}$$

称 $P(s,x,t,\cdot)$ 为**转移概率**，称方程(2.3)为 Chapman-Колмогоров **方程**. 1931 年 Колмогоров 研究了一类连续 Марков 过程，后来称之为**扩散过程**. 为简单起见，我们只考虑一维过程. Колмогоров 研究的扩散过程满足如下条件（$E[\cdot\,|X_t]$ 表示条件数学期望）：

$$E[X_{t+\triangle t}-X_t|X_t=x]=b(t,x)\triangle t+o(\triangle t), \tag{2.4}$$

$$E[(X_{t+\triangle t}-X_t)^2|X_t=x]=a(t,x)\triangle t+o(\triangle t), \tag{2.5}$$

称 $b(t,x)$ 为**漂移系数**，$a(t,x)$ 为**扩散系数**. 假定过程转移概率 $P(s,x,t,\cdot)$ 有密度函数 $p(s,x,t,y)$，并假定 $p(s,x,t,y)$ 有适当的光滑性，Колмогоров 证明了 p 满足如下两个方程（分别称为**后向和前向方程**）：

$$\frac{\partial}{\partial s}p(s,x,t,y)=-L_{s,x}p(s,x,t,y), \tag{2.6}$$

$$\frac{\partial}{\partial_t}p(s,x,t,y)=-\frac{\partial}{\partial y}[b(t,y)p(s,x,t,y)] + \frac{\partial^2}{\partial y^2}[a(t,y)p(s,x,t,y)], \tag{2.7}$$

其中

$$L_{s,x}=\frac{1}{2}a(s,x)\frac{\partial^2}{\partial x^2}+b(s,x)\frac{\partial}{\partial \lambda}. \tag{2.8}$$

方程(2.7)亦称为 **Fokker-Planck** 方程.

鞅(Martingale)是1939年J. Ville首先引进概率论的，他借用了法文 Martingale 有"倍赌策略"（即赌输后加倍赌注）这一含义. 该词的另一含义才是现在的中译名"鞅"（马颌韁）. 设 $\{X_n,\ n\geqslant 1\}$ 为一列可积随机变量. 令 $\mathscr{F}_n$ 表示使 $X_1,\cdots,X_n$ 为可测的最小 σ-域（即 $\mathscr{F}_n$ 由 $X_1,\cdots,X_n$ 生成，记为 $\mathscr{F}_n=\sigma(X_i,i\leqslant n)$）. 如果对每个 n，$E[X_{n+1}\,|\,\mathscr{F}_n]=X_n$，则称 (X_n) 为一**鞅**. 若把 X_n 设想为一赌徒在赌

完第 n 局后的赌金，则鞅性质反映了这种赌博是公平的. 对连续参数情形我们有相应的鞅的概念. 1953 年，Doob 在他的《随机过程》这部历史性专著中首次系统总结了他和 Lévy 等人有关鞅的理论及应用成果，使鞅论成了随机过程理论的一个独立分支.

Brown 运动(即 Wiener 过程)既是 Марков 过程又是鞅. 这决定了 Brown 运动在随机分析学乃至整个随机过程理论中处于一个中心地位. 下面我们将会看到，从 Brown 运动出发可以构造出一类扩散过程. 这就是下面要介绍的 Itô 随机积分和随机微分方程.

§3 Itô 随机分析及应用

§3.1 Wiener 积分和 O.-U. 过程 尽管 Brown 运动的轨道处处不可微，Wiener 还是定义了平方可积函数关于 Brown 运动的积分，设 (B_t) 为一维标准 Brown 运动，f 为 $(0,T]$ 上的阶梯函数，即

$$f(t) = \sum_{i=0}^{n-1} c_i I_{(t_i,t_{i+1}]}(t), \tag{3.1}$$

其中 $0 = t_0 < t_1 < \cdots < t_n = T$，则令

$$(f.B)_T = \int_0^T f(s)\mathrm{d}B_s = \sum_{i=1}^{n-1} c_i(B_{t_{i+1}} - B_{t_i}), \tag{3.2}$$

于是 $\boldsymbol{E}\,(f.B)_T = 0$，$\boldsymbol{E}\,(f.B)_T^2 = \int_0^T f(s)^2\mathrm{d}s$. 从而映射 $f \mapsto (f.B)_T$ 可以等距扩张为 $L^2((0,T])$ 到 $L^2(\Omega, \mathscr{F}, P)$ 中的线性映射. 我们称 $(f.B)_T$ 为 **Wiener 积分**，记为 $\int_0^T f(s)\mathrm{d}B_s$. 这是一均值为零、方差为 $\int_0^T f(s)^2\mathrm{d}s$ 的 Gauss 随机变量. 如果 T 变动，则我们得到一随机过程 $f.B$，它是 Wiener 不定积分.

1930 年 Ornstein 和 Uhlenbeck 建立了 Brown 运动的动力学

理论. 设 Brown 运动粒子在某一坐标方向的瞬时速度为$(V(t),t \geqslant 0)$,则由 Newton 定律,$V(t)$ 满足如下的 **Langevin 方程**:

$$\dot{V}(t) = -\beta V(t) + f(t), \tag{3.3}$$

其中 $-\beta V(t)$ 为摩擦力(阻力),$f(t)$ 为随机力. 随机力是周围液体分子对粒子随机碰撞产生的合力. 此方程的解为

$$V(t) = V_0 e^{-\beta t} + \int_0^t e^{-\beta(t-s)} f(s) \mathrm{d}s. \tag{3.4}$$

由统计物理知, 当 $t \to \infty$ 时 $V(t)$ 趋于 Maxwell 分布, 于是 $E|V(t)|^2 \to \frac{kT}{m}$. 由此推知 $f(t)$ 可用"**白噪声**"$\sqrt{\frac{2kT\beta}{m}}\dot{B}(t)$ 来近似,这里$\dot{B}(t)$ 是 $B(t)$ 的"微分",于是我们有

$$V(t) = V_0 e^{-\beta t} + \sqrt{\frac{2kT\beta}{m}} \int_0^t e^{-\beta(t-s)} \mathrm{d}B_s. \tag{3.5}$$

这一 Gauss 过程称为 **O.-U. 过程**. 这一 Brown 运动模型更接近于物理现实. 令 $X_t = X_0 + \int_0^t V(s)\mathrm{d}s$. 则$(X_t)$ 是 Brown 运动的位置过程. 由于 β 的数量级是 $10^8 \sim 10^{10}$,可以证明(X_t) 与参数为$\frac{2kT}{m\beta}$的 Wiener 过程很近似. 这从理论上说明了 Einstein 关于 Brown 运动的统计描述的正确性.

§3.2 Itô随机积分和随机微分方程 Itô 创立随机分析学的最初动机是通过 Brown 运动直接构造出扩散过程. 作为第一步,1944 年Itô 推广了 Wiener 积分. 定义了一类随机过程关于 Brown 运动的随机积分. 现简要介绍如下. 设(B_t) 为一维标准 Brown 运动. 令 $\mathscr{F}_t = \sigma(B_s, s \leqslant t)$. 设 H 为$(0,T]$ 上的一阶梯过程:

$$H_t = \sum_{i=0}^{n-1} \xi_i I_{(t_i, t_{i+1}]}(t), \tag{3.6}$$

其中 $0 = t_0 < t_1 < \cdots < t_n = T$,$\xi_i$ 为 $\mathscr{F}_{t_i}$可测的平方可积随机变

量. 令

$$(H.B)_T := \int_0^T H_s \mathrm{d}B_s = \sum_{i=0}^{n-1} \xi_i (B_{t_{i+1}} - B_{t_i}). \tag{3.7}$$

由于(B_t)为一个鞅,故易知$\mathbf{E}(H.B)_T = 0$,且有

$$\mathbf{E}(H.B)_T^2 = \sum_{i=0}^{n-1} \mathbf{E}[\xi_i^2](t_{i+1} - t_i) = \mathbf{E}\left[\int_0^t H_s^2 \mathrm{d}s\right]. \tag{3.8}$$

现令H为$(0,T]$上一随机过程,使得限于每个$(0,t]$,$0<t\leqslant T$,过程H作为(ω,t)的二元函数关于$\mathscr{F}_t \times \mathscr{B}((0,t])$可测,这样的过程称为**循序可测过程**. 在经典的文献中亦称为**非超前(non-anticipating)过程**. 利用等式(3.8)可以把上述积分推广到满足$\mathbf{E}\left[\int_0^T H_s^2 \mathrm{d}s\right] < \infty$的循序可测过程$H$. 称这样定义的积分为**Itô积分**. 如果$H$是$(0,\infty)$上的循序可测过程,使得对一切$t \in \mathbf{R}_+$,$\mathbf{E}\left[\int_0^t H_s^2 \mathrm{d}s\right] < \infty$,则随机不定积分$((H.B)_t)$为一个鞅(关于$(\mathscr{F}_t)$),且有轨道连续的修正.

设(X_t)为一扩散过程,满足(2.4)及(2.5). 1948年Lévy将过程(X_t)在短时间内的增量表达为

$$X_{t+\triangle t} - X_t = b(t,X_t)\triangle t + \sqrt{a(t,X_t)}\xi_t \sqrt{\triangle t} + o(\triangle t), \tag{3.9}$$

其中ξ_t为标准正态随机变量. Itô发现可用Brown运动的增量$B_{t+\triangle t} - B_t$代替$\xi_t\sqrt{\triangle t}$,然后令$\triangle t \to 0$便得到如下的"**随机微分方程**":

$$\mathrm{d}X_t = b(t,X_t)\mathrm{d}t + \sqrt{a(t,X_t)}\mathrm{d}B_t. \tag{3.10}$$

该方程可用Itô积分改写为**随机积分方程**:

$$X_t = X_0 + \int_0^t b(s,X_s)\mathrm{d}s + \int_0^t \sqrt{a(s,X_s)}\mathrm{d}B_s. \tag{3.11}$$

1951 年 Itô 采用 Picard 逐次逼近法证明了：如果对任给 $T>0$，$b(t,x)$ 及 $\sqrt{a(t,x)}$ 在 $[0,T]\times R$ 上关于 x 满足一致 Lipschitz 条件及线性增长条件，则方程(3.10)有唯一解，其中初值 X_0 是任意给定的与布朗运动 (B_t) 独立的平方可积随机变量. 这样，Itô 实现了通过布朗运动构造扩散过程的设想. 从此，一门崭新的概率论分支——**随机分析学**诞生了. 应该指出，Гихман 于 1947 年和 1950 年也独立地研究了不同形式的随机微分方程. 与 Колмогоров 及 Feller 用分析方法研究扩散过程相比，Itô 的**概率方法**的最大优点是直观、简便. 过程如何依赖于漂移系数及扩散系数从方程(3.10)一目了然. 此外，对这两个系数的光滑性要求比较弱. **随机微分方程(3.10)正是对扩散现象的微观概率机制的描述**. 从这一意义上说，Itô 的随机分析是随机王国中的 Newton 定律.

设 (H_t) 为关于 Brown 运动 (B_t) 可积的循序可测过程，(K_t) 为 Lebesgue 可积的**适应可测过程**(这里适应性是指对每个 t，K_t 关于 $\mathscr{F}_t$ 可测). 令

$$X_t = X_0 + \int_0^t H_s \mathrm{d}B_s + \int_0^t K_s \mathrm{d}s, \tag{3.12}$$

通常称这样的过程 (X_t) 为 **Itô 过程**. 关于 Itô 过程有如下的变量替换公式(称为 **Itô 公式**)：

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)(H_s \mathrm{d}B_s + K_s \mathrm{d}s) + \frac{1}{2}\int_0^t f''(X_s) H_s^2 \mathrm{d}s, \tag{3.13}$$

其中 f 为二次连续可微函数. 这一公式是随机分析的一个重要工具.

关于随机微分方程的研究有大量的文献. 其中一项重要的进展是 1969 年 Stroock 和 Varadhan 提出的研究方程弱解的"**鞅问题**"**方法**. 为叙述方便起见，我们考虑如下的**时齐随机微分方程**：

$$\mathrm{d}X_t^i = b^i(X_t)\mathrm{d}t + \sum_{\alpha=1}^{d} \sigma_\alpha^i(X_t)\alpha B_t^\alpha, 1 \leqslant i \leqslant d. \tag{3.14}$$

其中$(B_t^1,\cdots,B_t^d)$为d-维 Brown 运动. 给定(b^i)及(σ_α^i)，如果存在某概率空间$(\Omega,\mathscr{F},P)$及其上的 Brown 运动(B_t)及过程(X_t)使得(3.14)成立，则称(X,B)为(3.14)的一个**弱解**. "鞅问题"的基本出发点是：设$A=\frac{1}{2}\sum_{i,j=1}^{d}a^{ij}(x)\frac{\partial^2}{\partial x_i\partial x_j}+\sum_{j=1}^{d}b^j(x)\frac{\partial}{\partial x_j}$，其中$a^{ij}(x)=\sum_{\alpha,\beta}\sigma_\alpha^i(x)\sigma_\beta^j(x)$. 若$(X_t)$为方程(3.14)的一个弱解，则对每个$f\in C^\infty(R^d)$由 Itô 公式得

$$C_t^f\equiv f(X_t)-f(X_0)-\int_0^t Af(X_s)\mathrm{d}s$$
$$=\sum_{i=1}^{d}\sum_{\alpha=1}^{d}\int_0^t\frac{\partial}{\partial x_i}f(X_s)\sigma_\alpha^i(X_s)\mathrm{d}B_s^\alpha. \qquad (3.15)$$

于是(C_t^f)为一局部鞅(见 §3.3). 现令Ω为R_+上R^d-值连续函数全体，$\mathscr{F}$为Ω上的 Borelσ-域. 令$X_t(\omega)=\omega(t)$. 如果$(\Omega,\mathscr{F})$上的一族概率测度$(P_x,x\in R^d)$使得对每个$f\in C^\infty(R^d)$过程(C_t^f)在每个P_x下为局部鞅，则称$(P_x,x\in R^d)$是与算子A联系的"鞅问题"的一个解. 这时，"鞅问题"适定(即有唯一解)等价于对任意初始分布弱解存在且唯一(唯一性是指过程分布唯一). "鞅问题"方法不仅对构造和研究高维扩散及流形上的扩散很有用，而且在研究用扩散过程逼近其他过程时也非常有用.

§3.3 半鞅与随机分析 Itô 随机分析在 70 年代随着鞅论及随机过程一般理论的发展而形成为有关半鞅的随机分析理论. 在鞅论及随机过程一般理论中，我们通常取定一基本概率空间$(\Omega,\mathscr{F},P)$及$\mathscr{F}$的一族单调上升的子σ-域$(\mathscr{F}_t)$，并假定$\mathscr{F}$关于P是完备的，$\mathscr{F}_t$包含$\mathscr{F}$中一切零概集，而且假定$(\mathscr{F}_t)$是右连续的，即$\bigcap_{s>t}\mathscr{F}_s=\mathscr{F}_t$. 称这样的$\sigma$-域流$(\mathscr{F}_t)$满足**通常条件**.
一随机过程(X_t)称为$(\mathscr{F}_t)$鞅，如果(X_t)为$(\mathscr{F}_t)$-适应(即X_t为$\mathscr{F}_t$可测)，每个X_t可积，且对$s<t$有$E[X_t|\mathscr{F}_s]=X_s$. 设T为一非负随机变量. 如果对一切$t\geqslant 0$，$[T\leqslant t]\in\mathscr{F}_t$，则称$T$为停时.

设(X_t)为一$(\mathscr{F}_t)$-适应过程，如果存在**停时列**(T_n)使$T_n\uparrow\infty$，且使每个停止到T_n的过程$X^{T_n}(X_t^{T_n}(\omega)\equiv X_{t\wedge T_n(\omega)}(\omega))$为$(\mathscr{F}_t)$-鞅，则称$(X_t)$为$(\mathscr{F}_t)$-**局部鞅**. 一个适应过程如果可以表为一局部鞅与一适应的有限变差过程之和，就称此过程为**半鞅**. 这里我们定义的鞅、局部鞅及半鞅，都假定它们的轨道是**右连左极的**(即右连续且左极限存在). 半鞅是相当广泛的随机过程类，它包含一大类独立增量过程、Gauss 过程及 Марков 过程. 前一节中随机微分方程的解都是半鞅.

下面我们简要说明如何定义半鞅的随机积分. 首先，我们交代一些概念. 我们把由一切左连续$\mathscr{F}_t$适应的过程生成的$R_+\times\Omega$上的σ-域称为**可料σ-域**，关于此σ-域可测的过程称为**可料过程**. 设M为一连续局部鞅，则可以证明(利用鞅论中的 Doob-Meyer 分解定理)存在唯一的连续适应增过程(记为$\langle M,M\rangle$)使得$M^2-\langle M,M\rangle$为零初值局部鞅. 设(X_t)为一半鞅，则可以唯一确定它的"**连续局部鞅部分**"(X_t^c)，令$[X,X]_t=\langle X^c,X^c\rangle_t+\sum_{s\leqslant t}\triangle X_s^2$. 称增过程为半鞅$X$的**二次变差过程**. 对两个半鞅$X$、$Y$，自然可以定义**互变差过程**$[X,Y]$. 推广 Itô 随机积分的关键一步是定义可料过程关于平方可积鞅的随机积分. 这是由 H. Kunita 和 S. Watanabe 于 1967 年完成的. 设M为一平方可积鞅，$L_2(M)$为使$E\left[\int_{0-}^{\infty}H_t^2\mathrm{d}[M,M]\right]<\infty$的可料过程$H$的全体. 设$H\in L_2(M)$，则对任何平方可积鞅$N$，我们有如下的 Kunita-Watanabe 不等式

$$\left|E\int_{0-}^{\infty}H_t\mathrm{d}[M,M]_t\right|\leqslant\left(E\int_{0-}^{\infty}H_t^2\mathrm{d}[M,M]_t\right)^{\frac{1}{2}}\left(E[N,N]_\infty\right)^{\frac{1}{2}}. \tag{3.16}$$

于是存在唯一的平方可积鞅L，使得对任何平方积鞅N有$E[L,N]_\infty=E\left[\int_{0-}^{\infty}H_t\mathrm{d}[M,N]_t\right]$(Riesz 表现定理). 称$L$为$H$关于$M$的**随机(不定)积分**. 另一方面，如果一局部鞅M的轨道是有限变差

函数(即在有界区间上为有界变差的),且可料过程 H 的几乎所有轨道关于 M 的相应轨道的 Stieltjes 积分存在,则自然将这一按轨道的 Stieltjes 积分定义为随机积分. 可以证明,这样定义出的积分仍是一局部鞅. 在这里被积过程的可料性是关键. 由于任一局部鞅可以分解为一局部平方可积鞅与一有限变差局部鞅之和,所以不难定义可料过程 H 关于局部鞅的随机积分. 这时可积性的充要条件是 $\sqrt{\int_0^t H_s^2 \mathrm{d}[M,M]_s}$ 为局部可积增过程. 进一步,就可定义可料过程关于半鞅的随机积分. 随机积分的这一理论主要是由 P. A. Meyer 为首的斯特拉斯堡学派完成的. 已经证明:如果要求随机积分关于被积过程(甚至限于阶梯可料过程)有某种连续性,则半鞅是能够定义随机积分的最大过程类.

设 (X_t) 及 (Y_t) 为两个半鞅,我们有如下的**分部积分公式**:

$$X_tY_t = \int_0^t X_{s-}\mathrm{d}Y_s + \int_0^t Y_{s-}\mathrm{d}X_s + [X,Y]_t. \tag{3.17}$$

设 $(X_t^1),\cdots,(X_t^n)$ 为半鞅,$f(x_1,\cdots,x_n)$ 为 R^n 上的二次连续可微函数,则有如下的 **Itô 公式**:

$$\begin{aligned} & f(X_t^1,\cdots,X_t^n) \\ & = f(X_0^1,\cdots,X_0^n) + \sum_{i=1}^n \int_0^t D_i f(X_{s-}^1,\cdots,X_{s-}^n)\mathrm{d}X_s^i \\ & \quad + \frac{1}{2}\sum_{i,j=1}^n \int_0^t D_iD_j f(X_{s-}^1,\cdots,X_{s-}^n)\mathrm{d}\langle (X^i)^c,(X^j)^c\rangle_s \\ & \quad + \sum_{0<s\leq t}\Big[f(X_s^1,\cdots,X_s^n) - f(X_{s-}^1,\cdots,X_{s-}^n) \\ & \qquad - \sum_{i=1}^n D_i f(X_{s-}^1,\cdots,X_{s-}^n)\triangle X_s^i\Big]. \end{aligned} \tag{3.18}$$

设 (X_t) 为一半鞅,令

$$\mu(\mathrm{d}t,\mathrm{d}x) = \sum_{s>0} I_{[\Delta x_s\neq 0]}\,\mathscr{E}_{[s,\Delta X_s]}(\mathrm{d}t,\mathrm{d}x), \tag{3.19}$$

其中 $\mathscr{E}(s,y)$ 为只在 (s,y) 处有单位负荷的测度,称 μ 为与 X 联系

的随机测度. 则 X 有如下典则表示：

$$X_t = \alpha_t + X_t^c + \int_0^t \int_{|x|\leq 1} x\mathrm{d}(\mu - \nu) + \int_0^t \int_{|x|>1} x\mathrm{d}\mu. \tag{3.20}$$

其中 ν 为 μ 的**可料对偶投影**,(α_t) 为一可料有限变差过程. 令 $\beta = \langle X^c, X^c \rangle$. 称$(\alpha,\beta,\nu)$ 为半鞅(X_t) 的**局部特征**. 这一概念是从 Lévy 关 于独立增量过程的分解定理引伸出来的. 虽然半鞅的局部特征并不能完全刻画半鞅本身,但许多有关半鞅的结果(如半鞅的泛函中心极限定理)可以用半鞅的局部特征表述.

§3.4 Itô 随机分析的应用 Itô 随机分析有广泛的应用. 我们举某些有代表性的例子简要介绍之.

(1) 偏微分方程解的概率表示 设 L 为 R^d 上的二阶椭圆算子：

$$L = \frac{1}{2}\sum_{i,j=1}^{d} a^{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{j=1}^{d} b^j(x)\frac{\partial}{\partial x_j}. \tag{3.21}$$

设 C 为 R^d 上的有界连续函数,考虑 Cauchy 问题：

$$\frac{\partial u}{\partial t} = Lu + C(x)u,\ u(0,x) = f(x). \tag{3.22}$$

设与 L 联系的鞅问题是适定的,$(P_x, x\in R^d)$ 为鞅问题的唯一解,则(3.22) 的解由如下的 **Feynman-Kac** 公式给出：

$$u(t,x) = E_x\left[f(X_t)\exp\int_0^t C(X_s)\mathrm{d}s\right], \tag{3.23}$$

其中(X_t) 为连续函数空间 $C(R_+\to R^d)$ 上的典则过程：$X_t(\omega) = \omega(t)$. 这一概率表示的优点是：(1) 对算子 L 的系数光滑性要求较弱；(2) 可以通过 L- 扩散过程的性质推出方程解的性质；(3) 可用 Monte-Carlo 方法求数值近似解.

(2) 随机最优控制 设 Ω 为 R_+上取值于 R^d 的连续函数全体. $\mathscr{F}$ 为 Ω 的 Borel σ- 域, P 为$(\Omega, \mathscr{F})$ 上的概率测度,使得坐标过程 $B_t(\omega) = \omega(t)$ 为 d- 维标准 Brown 运动. 在一定条件下,如下方程有唯一解：

$$\mathrm{d}X_t = \sigma(t,X)\mathrm{d}B_t,\ 0 \leqslant t \leqslant T. \tag{3.24}$$

一受控动态系统通常由下述随机微分方程来描述：

$$\mathrm{d}Y_t = f(t,Y,u_t)\mathrm{d}t + \sigma(t,Y_t)\mathrm{d}B_t, 0 \leqslant t \leqslant T. \tag{3.25}$$

其中(u_t)为取值于某距离空间的控制过程. 在对f及σ作某些假定下，对每个容许控制u，在$(\Omega, \mathscr{F})$上存在一与P等价的概率测度P^u，使得过程(X_t)为方程(3.25)在$(\Omega, \mathscr{F}, P^u)$上的一个弱解. 于是，控制的效应体现为轨道空间$\Omega$上概率测度的等价改变. 从这种观点出发，一个重要的结果是：与容许控制u联系于一**代价过程**(M_t^u)，它在P^u下为下鞅，并且当且仅当u为**最优控制过程**时，(M_t^u)在P^u下为鞅.

(3) **非线性滤波**　信号过程(X_t)及观测过程(Y_t)通常用如下两个随机微分方程描述：

$$\mathrm{d}X_t = g(X_t)\mathrm{d}t + \sigma(x_t)\mathrm{d}B_t,\ 0 \leqslant t \leqslant T. \tag{3.26}$$

$$\mathrm{d}Y_t = h(X_t)\mathrm{d}t + \alpha(Y_t)\mathrm{d}W_t. \tag{3.27}$$

这里(B_t)和(W_t)分别为d-维和m-维Brown运动. 设$\mathscr{Y}_t = \sigma(Y_s, s \leqslant t)$，$\varphi$是$R^d$上一有界可测函数，则通常所谓的滤波问题是要计算条件期望$E[\varphi(X_t) \mid \mathscr{Y}_t]$. 它是已知观测$(Y_s, s \leqslant t)$对$\varphi(X_t)$所作的最好估计. 问题归结为求$X_t$关于$\mathscr{Y}_t$的条件分布. 假定其条件分布密度$\hat{p}(t,x)$存在，在对方程(3.26)及(3.27)的系数作适当假定下，可以证明$\hat{p}(t,x)$由如下公式给出：

$$\hat{p}(t,x) = q(t,x)\Big/\int_{R^d} q(t,y)\mathrm{d}y. \tag{3.28}$$

而$q(t,x)$是如下**随机偏微方程**的解：

$$\mathrm{d}q(t,x) = L^* q(t,x)\mathrm{d}t + h(x)q(t,x)\mathrm{d}Y_t, \tag{3.29}$$

其中L^*为L的共轭算子，$L = \frac{1}{2}\sum_{i,j} a^{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_j g^j(x)\frac{\partial}{\partial x_j}$，$(a^{ij}) = (\sigma_\alpha^i)(\sigma_\alpha^j)^\tau$. 方程(3.29)称为**Zakai 方程**.

(4) **动态系统的小随机扰动**　设$\dot{X}(t) = b(X(t))$为一(确定

型）动态系统. 假定系统受到一白噪声型的小随机扰动，则系统由下述随机微分方程描述（$\epsilon \neq 0$ 为一小参数）：

$$dX_t^\epsilon = b(X_t^\epsilon) + \epsilon\sigma(X_t^\epsilon)dW_t, \tag{3.30}$$

其中(W_t) 为一 d- 维 Brown 运动. 即使原系统有渐近稳定的平衡点，小随机扰动的累积效应可以使系统概率为 1 地离开任意有界区域. 小随机扰动的动态系统有广泛的应用. 例如，对时间及空间尺度作某种改变后，化学反应中分子在势井中的运动可由小随机扰动的动态系统模拟，其化学反应速率与分子从势井中逃逸的平均时间成反比. 对随机系统(3.30)，人们关心的是下述各值按 ϵ 的渐近展开：(a) 首次离去某区域时离出点在边界上的概率分布；(b) 期望离出时间. 这两个问题可以归结为奇异摄动椭圆型边值问题. 后一问题与大偏差理论有联系（见 §4.4）.

(5) **随机过程的扩散逼近** 在许多实际问题中，我们建立的数学模型未必是扩散过程. 但是在一定条件下，通过时空尺度的改变，我们可以用扩散过程来逼近实际的随机过程. 一个典型的例子是群体发生数学模型，它通常用 Марков 链来描述. 如果我们想计算某物种灭绝的概率，当群体的个体数目很大时，这是很难做到的. 但是如用扩散过程来逼近尺度改变后的 Марков 链，则灭绝概率就相当于扩散过程从区间(0,1)中由边界点 0 流出的概率. 而后者是可以实际计算的.

§4 随机分析的若干研究方向

现在我们介绍随机分析的若干较新的研究方向. 由于篇幅所限，我们只作粗略的介绍.

§4.1 Malliavin 分析 设Ω为R_+上零初值连续函数全体，它按有界区间上一致收敛拓扑为一 Frechet 空间. 令 $\mathscr{B}(\Omega)$ 为其上的 Borel σ- 域. 则存在$(\Omega, \mathscr{B}(\Omega))$ 上唯一概率测度 P，使得坐

标过程 $W_t(\omega)=\omega(t)$ 为一标准 Wiener 过程. 称 P 为 **Wiener 测度**. $(\Omega, \mathscr{B}(\Omega), P)$ 上的可测函数称为 **Wiener 泛函**(或 **Brown 泛函**). 常见的 Wiener 泛函是Itô 积分或Itô 随机微分方程的解. 这些 Wiener 泛函一般说来关于 Ω 上的拓扑不连续,因此无法定义它们的微分. 1976年P. Malliavin以给出偏微分方程中著名的 Hörmander 次椭圆性定理的概率证明为动机,首次引进了一类 Wiener 泛函的"相对微分"运算,从而开创了一种崭新的随机分析学. 现在通常称为 **Malliavin 分析**. 由于 Malliavin 引进的 Wiener 泛函的微分是通过 Ω 中元素的"变分"来定义的,所以 Malliavin 分析亦称为 **Malliavin 随机变分**. 这里"变分"不是传统意义下使泛函达到极小值,而是"扰动"的意思. 具体来说,设 H 为 R_+ 上绝对连续且其导数平方可积的函数全体. 称 H 为 **Cameron-Martin 空间**. 设 $h\in H$,令 $\tau_h\omega(t)=\omega(t+h)$,则 P 在 τ_h 下的象测度 $P\circ\tau_h^{-1}$ 与 P 相互绝对连续. 若 $h\notin H$,则 $P\circ\tau_h^{-1}$ 与 P 相互奇异. Wiener 泛函 $F(\omega)$ 称为在 Malliavin 意义下可微,是指存在$(\Omega, \mathscr{F}, P)$ 上一 H- 泛函,记为 $DF(\omega)$,使得

$$\frac{\mathrm{d}}{\mathrm{d}t}F(\omega+th)|_{t=0}=\langle DF(\omega),h\rangle_H, \quad \forall h\in H. \quad (4.1)$$

这里 $\omega+th$ 就是 ω 的"变分"(variation). D 称为**梯度算子**. 其共轭算子 δ 称为**散度算子**. Malliavin 分析最初讨论的中心问题是:(1) 何时 Wiener 泛函 F 的分布关于 Lebesgue 测度绝对连续?(2) 如果绝对连续,进一步问密度的光滑性如何?这些问题都已彻底解决. 在实际应用中,关键是要计算 **Malliavin 协方差矩阵** $\langle DF,DF\rangle$. (当 F 是 R^d- 值泛函时,这是 $d\times d$ 矩阵.) 例如,当 F 是一 Itô 随机微分方程解时,可以求得它的表达式. 如果 Wiener 泛函 F 在 Malliavin 意义下是无穷次可微的,且$[\det\langle DF,DF\rangle]^{-1}\in L^p, \forall p>1$,则 F 的分布关于 Lebesgue 测度绝对连续,且其密度函数是 C^∞ 的. 这里关于$\langle DF,DF\rangle$ 的条件称为 **在 Malliavin 意义下非退**

化.

Malliavin分析是Wiener空间上的微分学. 这是一个新的理论大厦. 通向这一大厦的途径有三条：一是由 Malliavin 首创并由 Stroock 发展的无穷维 O.-U. 过程无穷小算子方法；二是 Bismut 的 Girsanov 变换方法；三是由 Shigekawa 提出并由 Meyer 及 Watanabe 等发展的泛函分析方法. 按照第三种方法，Malliavin 分析就是 Wiener 泛函的 Соболев 空间理论. 这样，我们就可以研究一类 **广义 Wiener 泛函** . Watanabe 证明了 Schwartz 缓增广义函数与一在 Malliavin 意义下非退化的光滑 Wiener 泛函的“复合”是一广义 Wiener 泛函. 这时，$p(x) = \langle \delta_x(F), 1 \rangle$ 正是 F 的分布密度.

Malliavin 分析有许多重要应用. 其中最引人注目的是 Malliavin 关于 Hörmander 定理的概率证明和 Bismut 关于 Atiyah-Singer 指标定理的概率证明. 此外，Malliavin 分析在非线性滤波、热核(heat kernel) 小时间渐近估计、随机振盪积分估计、大偏差等问题中获得了成功的应用. 现在的主要问题是进一步扩大它的应用领域. 新近发展起来的“拟必然分析”是一个值得注意的新方向.

§ 4.2 Dirichlet 型与对称 Марков **过程** 1959 年 Beurling 和 Deny 首次引进 Dirichlet 空间的概念，其目的是建立位势理论的 L^2- 框架. 在 70 年代，主要由 Fukushima 及 Silverstein 把它进一步发展成为一个系统的 Dirichlet 型理论. 这一理论是联系位势论和 Марков过程的一个桥梁，由于它在非相对论量子力学及 Euclid 场论中有重要应用. 近年来这一理论有了迅速发展. 设 E 为一可分距离空间，m 为 $\mathscr{B}(E)$ 上的一σ-有限测度，则 $H = L^2(E, m)$ 为一可分 Hilbert 空间，设$(\mathscr{E}, \mathscr{D})$为 $L^2(E, m)$ 上的一非负对称稠定双线性型. 令 $\mathscr{E}_1(u, v) = \mathscr{E}(u, v) + (u, v), u, v \in \mathscr{D}$. 若 $\mathscr{D}$ 关于 ε_1- 范数完备，则称$(\varepsilon, \mathscr{D})$为闭的. 如果

$$u \in D \Rightarrow u^+ \wedge 1 \in \mathscr{D}, \text{且} \mathscr{E}(u^+ \wedge 1, u^+ \wedge 1) \leqslant \mathscr{E}(u, u),$$

则称($\mathscr{E}$,$\mathscr{D}$)有**压缩性**(或 Марков 性). 我们把一具有 Марков 性的非负对称闭双线性型($\mathscr{E}$,$\mathscr{D}$)称为 Dirichlet 型. Dirichlet 型与 H 上强连续对称 Марков 半群(T_t)有一一对应关系:

$$A \leftrightarrow \mathscr{E}(u,v) = (\sqrt{-A}u, \sqrt{-A}v), \mathscr{D}(\mathscr{E}) = \mathscr{D}(\sqrt{-A}).$$

其中 A 为(T_t)的生成算子. 如果 E 为局部紧第二可数空间,且 Dirichlet 型($\mathscr{E}$,$\mathscr{D}(\mathscr{E})$)为正则的(即 $C_0(E) \cap D(\mathscr{E})$在 $\mathscr{D}(\mathscr{E})$中按 $\mathscr{E}_1$- 范数稠,在 $C_0(X)$ 中按一致范数稠),则 Fukushima 证明了半群(T_t)为一 m- 对称的 Hunt 过程的 L^2- 半群. 新近,S. Albeverio 和 Z. M. Ma(马志明)在取消空间的局部紧性及 Dirichlet 型的正则性这两个限制条件下,建立了一类广泛的拟正则 Dirichlet 型与一类 m- 对称的 Марков 过程之间的一一对应关系. 这一重要结果特别适用于无穷维空间中的 Dirichlet 型,从而对无穷维分析和 Euclid 量子场有重要意义.

下面举例说明 Dirichlet 型对量子物理的应用. 在非相对论量子力学中,一个具有 d- 个自由度的系统的动力学由 $L^2(R^d, \mathrm{d}x)$ 上的一非负自共轭算子 H 决定:$H = -\frac{1}{2}\triangle + V$,其中 $\triangle$ 为 Laplace 算子,V 为位势函数. H 称为**能量算子**或 **Hamilton 算子**. 假定算子 H 关于 $C_0^\infty(R^d)$ 为本质自伴的且下有界的,则 H 的最小谱为单重的,相应的本征函数 φ 可取为几乎处处严格正,且$\int \varphi^2 \mathrm{d}x = 1$. 称 φ 为**基态**(或**真空态**). 为简单起见,不始设 H 的最小谱为 0. 令 $\nu(\mathrm{d}x) = \varphi^2(x)\mathrm{d}x$, $Wf(x) = f(x)/\varphi(x)$,则 W 为 $L^2(R^d, \mathrm{d}x)$ 到 $L^2(R^d, \mathrm{d}\nu)$ 上的酉变换. 令 $H_\nu = WHW^{-1}$. 则 H_ν 为 $L^2(R^d, \nu)$ 上的非负自共轭算子,称 H_ν 为**基态表示**. 如果 φ 比较光滑,则由分部积分公式知 $H_\nu = -\frac{1}{2}\triangle + b\nabla$,其中 $b = -\nabla\ln\varphi$. 但当 V 是奇异位势时,H_ν 没有上述表达式. 这时令

$$\varepsilon_\nu(u,v) = \frac{1}{2}\int_{R^d} \nabla u \cdot \nabla v \, \nu(\mathrm{d}x), u, v \in C_0^\infty(R^d).$$

则($\mathscr{E}^{\nu}$, $C_o^\infty(R^d)$)的最小闭扩张为一 Dirichlet 型(称为经典 Dirichlet 型). 它联系于一扩散过程. 用这种基态表示,我们可以定义和讨论具有奇异位势 V 的 Hamilton 算子. 在量子场论中,与之相应的是无穷维经典 Dirichlet 型

$$\mathscr{E}(f,f)=\int_{\mathscr{S}'(R^d)} |\nabla f|^2 \mathrm{d}\nu,$$

其中 $\mathscr{S}'(R^d)$ 为 Schwartz 缓增广义函数空间,ν 为一概率测度.

§4.3 白噪声分析 白噪声分析(white noise analysis)是 1975 年由飞田武幸(T. Hida)首先提出的. 其基本思想是把 Brown 泛函 $f(B(t), t\in R)$ 视为白噪声($\dot{B}(t), t\in R$)的泛函. 具体来说,设 $\mathscr{S}'(R)$ 为 Schwartz 缓增广义函数空间,$\mathscr{B}$ 为其 Borel σ-域,则由 Minlos 定理,存在($\mathscr{S}'(R)$, $\mathscr{B}$)上唯一的概率测度 μ,使得

$$\int_{\mathscr{S}'(R)} e^{i\langle x,\xi\rangle}\mu(\mathrm{d}x)=\exp\left\{-\frac{1}{2}|\xi|_2^2\right\}, \xi\in\mathscr{S}(R).$$

基中 $|\cdot|_2$ 为 $L^2(R)$ 的范数,$\mathscr{S}(R)$ Schwartz 速降 C^∞ 函数空间.

称 μ 为**白噪声测度**. 如果我们将白噪声过程($\dot{B}(t), t\in R$)视为 $\mathscr{S}'(R)$-值随机元,则测度 μ 恰好是它的分布. ($\mathscr{S}'(R)$, $\mathscr{B}$, μ)上的可测函数称为**白噪声泛函**. 显然,Brown 泛函可以表示为白噪声泛函. 另一方面,我们可在**白噪声空间** ($\mathscr{S}'(R)$, $\mathscr{B}$, μ)上构造一 Brown 运动(B_t):

$$B_t=\begin{cases}\langle\cdot, I_{(0,t]}\rangle, t>0,\\ \langle\cdot, I_{[t,0]}\rangle, t\leqslant 0,\end{cases}$$

其中 $\langle\cdot, I_{[0,t]}\rangle$ 表示 $\langle\cdot, f^{(n)}\rangle$ 的 L^2- 极限,$(f^{(n)})\subset\mathscr{S}(R)$,且$(f^{(n)})$ 在 $L^2(R)$ 中趋于 $I_{[0,t]}$. 将 Brown 泛函视为白噪声泛函的最大优点是可以充分利用空间 $\mathscr{S}'(R)$ 的线性拓扑结构($\mathscr{S}'(R)$ 是核空间 $\mathscr{S}(R)$ 的对偶. 例如,设 $x\in\mathscr{S}'(R)$. 我们可以定义张量乘积 $x^{\otimes n}$(它属于 $\mathscr{S}'(R_1^n)$))的 **Wick 变换** $:x^{\otimes n}:$. 如果 f 是 R^n 上的对称

函数，且属于 p 阶 Соболев 空间，$p>\frac{1}{2}$，则 f 关于 Brown 运动 (B_t) 的 **n-重 Wiener 积分** $I_n(f)(x)$ 可以表示为 $\langle :x^{\otimes n}:, f\rangle$. 这一表达式给研究 Brown 泛函（即白噪声泛函）带来很大方便. 此外，用**二次量子化**方法，可以从 R 上的 Соболев 空间 $\mathscr{S}_p(R)$ 出发建立白噪声空间 $(\mathscr{S}'(R),\mu)$ 上的 **白噪声泛函的** Соболев **空间** $(\mathscr{S})_p$. 令

$$(\mathscr{S})=\bigcap_{p\in R_+}(\mathscr{S})_p,\quad(\mathscr{S})^*=\bigcup_{p\in R_+}(\mathscr{S})_{-p},$$

称 $(\mathscr{S})$ 为 Hida **检验泛函空间**，称 $(\mathscr{S})^*$ 为 Hida **广义泛函空间**. Hida 意义下的广义泛函比前面（§4.1）提到的 Watanabe 意义下的广义泛函更广泛. 例如，白噪声 $\dot{B}(t)$ 是 Hida 广义泛函，但不是 Watanabe 意义下的广义泛函.

白噪声分析在量子力学、随机分析、Dirichlet 型及量子场论中获得了成功应用. 例如，量子力学中的 Feynman 积分中的形式"测度"可以解释为 Hida 意义下的复值广义泛函；$P(\varphi)_2$ 场的场测度是正的 Hida 广义泛函；Itô 随机积分 $\int_0^t H_s \mathrm{d}B_s$ 可以理解为 Bochner 积分 $\int_0^t H_s\dot{B}(s)\mathrm{d}s$. 白噪声分析的理论框架刚刚建成不久，因此可望今后将在理论及应用方面都有新的发展.

§4.4　大偏差理论　设 E 是一 Polish 空间，$\{\mu_\varepsilon,\varepsilon>0\}$ 是其上一族概率测度，使得当 $\varepsilon\to 0$ 时 μ_ε 弱收敛于 δ_{x_0}，设 $\Gamma\in\mathscr{B}(E)$，使 $x_0\notin\overline{\Gamma}$，则 $\lim\limits_{\varepsilon\to 0}\mu_\varepsilon(\Gamma)=0$. 大偏差理论是要研究 $\mu_\varepsilon(\Gamma)$ 趋于 0 的速度. 大偏差理论源于大样本统计. 我们希望从收敛于极限分布的速度来决定取样大小. 历史上最早的大偏差结果是 Cramér 于 1938 年得到的. 他证明了如下结果：若母体分布的母函数 $M(\theta)=E[e^{\theta X}]<\infty\ \forall\ \theta\in R$，则独立采样的样本均值与母体均值偏离大于某个正数的概率按指数速度趋于零. 60 — 70 年代 Varadhan 和 Donsker 在前人工作的基础上创立了大偏差理论的一般框架并建

立了有关 Марков 过程大偏差的系统的结果.

设 E 为一 Polish 空间,函数 $I(x):E\mapsto[0,\infty]$ 称为**速率函数**,如果 I 不恒为 $+\infty$,且对一切 $l\geqslant 0$,$\{x:I(x)\leqslant l\}$ 是紧集. $(E,\mathscr{B}(E))$ 上的一概率测度族 $\{\mu_\varepsilon,\varepsilon>0\}$ 称为满足大偏差原理,且速率为 I,是指

$$\overline{\lim_{\varepsilon\downarrow 0}}\varepsilon\log\mu_\varepsilon(F)\leqslant-\inf_{x\in F}I(x),\ \forall\ \text{闭集}\ F,$$

$$\underline{\lim_{\varepsilon\downarrow 0}}\varepsilon\log\mu_\varepsilon(G)\geqslant-\inf_{x\in G}I(x),\forall\ \text{开集}\ G.$$

大偏差原理的最大用处是处理 Laplace 积分的渐近估计. Varadhan 的"**骨架定理**"是:如果$(\mu_\varepsilon,\varepsilon>0)$ 满足大偏差原理,且其速率函数为 I,则对 E 上任何有界连续函数 Φ,有

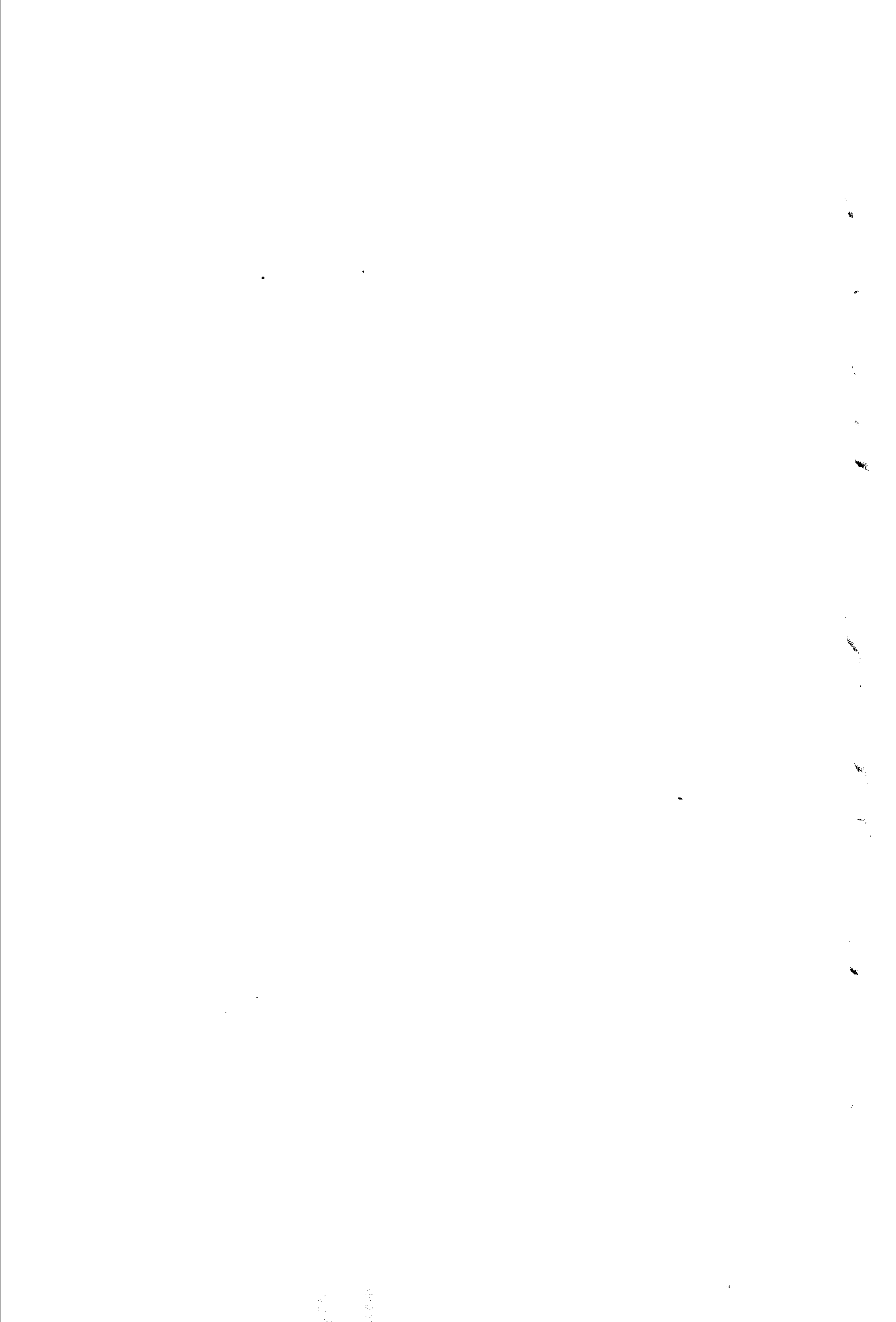
$$\lim_{\varepsilon\to 0}\varepsilon\log\int_E e^{\Phi/\varepsilon}\mathrm{d}\mu_\varepsilon=\sup_{x\in E}(\Phi(x)-I(x)).$$

一个典型的例子是:μ_ε 是与某二阶椭圆算子 L_ε 相应的扩散过程的分布,其中 ε 是小参数,Φ 是关于过程的轨道的泛函(例如泛函积分). 这种 Laplace 积分的渐近估计在统计物理及随机动态系统中有重要应用.

§4.5 其他方向 属于随机分析及相关领域的研究方向还有随机偏微分方程、无穷维随机微分方程、随机微分几何、流形上的随机分析、量子随机分析及随机量子化等. 此外,近年来由研究期权定价、消费-投资问题、市场的完备性、投资组合及套期保值等金融问题发展起来的数理金融学(Mathematical Finance)以及由分析和设计排队网络、通讯和计算机网络等实际需要而发展起来的"随机网络"(Stochastic Network)是应用随机分析学的两个新领域,吸引了众多的随机分析专家从事这两方面的研究. 因文章篇幅所限,就不在此一一介绍了.

参 考 文 献

1. 黄志远:随机分析学基础,武汉大学出版社,1988.
2. 严加安:鞅与随机积分引论,上海科技出版社,1981.
3. 龚光鲁:随机微分方程引论,北京大学出版社,1987.
4. Zeev Schuss:随机微分方程理论及其应用(中译本),上海科技文献出版社,1986.
5. J.-D. Deuschel and D. W. Stroock: Large Deviations, Pure and Appl. Math. Vol. 139, Academic Press. 1989.
6. M. Fukushima: Dirichlet Forms and Markov Processes, Tokyo, Kodansha and North-Holland, 1980.
7. S. W. He, J. G. Wang and J. A. Yan: Semimartingale Theory and Stochastic Calculus, Science Press and CRC Press, 1992.
8. I. Karatzas and S. E. Shreve: Brownian Motion and Stochastic Calculus, Springer-Verlag, 1987.
9. Z. M. Ma and M. Röckner: An introduction to the Theory of (Non-symmetric) Dirichlet Forms, Springer-Verlag, 1992.



X

多维数据分析与统计推断
Multidimensional Data Analysis and Statistical Inference

陈希孺
(中国科学技术大学研究生院)
Chen Xiru
(Graduate School of University
of Science and Technology of China)
成　平
(中国科学院系统科学研究所)
Cheng Ping
(Institute of Systems Science,
Academia Sinica)

高维数据统计分析及其他
Statistical Analysis of Multidimensional Data and Related Topics

中国数学发展的若干主攻方向
Some Main Aspects on Development
of Mathematics in China

一 引 言

严格说来,"高维数据统计分析"并不是一个新的课题,而是随着统计学的发展"与生俱来"的.例如,相关分析中要考虑多个相互有影响的变量,古典的线性回归分析以及基於多维正态分布的多元统计分析,原则上对所考察的变量数目也并无硬性的限制.即使在计算机应用尚不普及的时代,在实用问题中涉及多至数十个变量的情况,也并非绝无仅有.

但是近二十年左右以来,随着以下诸多方面的发展,给问题注入了一些新的因素.一是许多大型的问题涉及为数极大的变量.比如在关于环境污染之类的问题中,有可能涉及数以百计的变量.这类问题如用传统方法处理,不仅计算量过大,结果也不理想.二是高性能计算机的广泛应用,为处理涉及大量变量的问题提供了必要的计算手段,为高维数据统计分析的广泛应用提供了可能,并刺激了理论及方法上的研究工作的进展.三是一批新概念和新方法的提出,它们在不同程度上打破了传统方法的思维定势,给这门学科注入了新的活力,而赋予它一种新的面貌.例如,传统的多元分析着重多元正态分布,但在涉及大量变量的情况下,正态性假定往往过于勉强,与实际情况相去较远.因而迫切需要建立有关的理论和方法去处理非正态数据的问题.在这方面学者们作了不少努力,并已取得了一些有价值的成果.又如,一般说来,传统的统计分析方法比较着重分析单个(或若干)变量(即因子)的作用.在涉及很大数目的变量的统计分析问题中,单个或少数变量起的作用一般不大.更重要的是一种"整体"概念,即用尽可能简单和有效的方式,把包含在很复杂的数据中的信息尽可能提出来并给予一种富有统计意义的解释.下文将要介绍的投影寻踪方法,即是朝着这个方向的努力的一个重要例子.

由于以上这些发展，使高维数据的统计分析成为一个有活力的、前沿性的统计研究方向，有着重大的实用和理论上的重要性.

高维数据分析的困难可以从另一角度去考察. 假定有一个只含一个自变量 X_1 的回归问题，X_1 的变化范围为[0,1]. 把这区间分成 100 等分，在每个等分内选定一个 X_1 值进行观察，则大体上可以设想：所得数据可以对模型的局部变化弄清到 0.01 这样的细度. 但是，如果有 10 个自变量 $X_1,\cdots,X_{10}$，而想要把模型的局部变化情况明确到同样的细度，即每个边长为 0.01 的 10 维正立方体中都有 $(X_1,\cdots,X_{10})$ 的一个选定值，则至少需要有 10^{10} 这么多数据. 哪怕你有 10,000 个数据，其空间分布仍会显得十分稀疏. 又如设想有 n 维变量 $(X_1,\cdots,X_n)$ 的若干数据在体积的意义下均匀分布于 n 维欧氏空间 R^n 的单位球内. 则在离球的表面不到 0.1 的距离内，包含了全部数据总数的百分之 $100(1-(0.9)^n)$. 当 n 较大时此数接近 100，即实际上绝大多数数据都汇集在球面附近. 这种现象的存在使传统的统计分析方法难于得出满意的结果. 这种现象也正是华罗庚教授和王元教授在使用数论方法作近似高维积分时考虑的出发点. 事实证明，华 - 王的方法对高维数据统计分析也有用.

国内这方面的研究工作可以说始于 80 年代中期. 在国家自然科学基金委的支持下，国内若干单位的一些研究工作者集合在“高维数据统计分析的理论和方法”这个课题组的名目下，分头进行了研究工作并取得了一定的成果. 本文就是有关情况的一个概括的介绍. 我们也希望，本文能起到某种普及性的作用，即对于那些关心统计学的理论进展及其应用，但并不以本文所论题目为研究方向的读者，可以通过本文了解到有关论题的一些基本统计思想、发展概况及其在应用上的意义.

需要说明的是，本文所论及的某些内容，从狭义的观点看通常并不划入高维数据统计分析这个范围. 例如最小 L_1 模估计和 M 估

计方法，是近年发展起来的、在理论和应用上都受到重视的参数估计方法，它可以用于涉及大量自变量的回归问题，在不少情况下与传统的最小二乘法比较有其优越性，但在统计学中这个题目一般都被列入参数估计范围. 又如随机加权法，源出于 Bootstrap 方法，就其性质可以归入“数据分析”之内，但这种方法并非专为针对高维数的数据而设. 本文的内容，系由我们所从事的基金课题所决定. 但也应当看到：不应从过于狭义的观点来评估哪些内容应否属於“高维数据统计分析”的范围. 例如，上述从狭义的观点可以认为 Bootstrap 不属于高维数据统计分析的范畴. 但是，例如在一种典型的高维数据分析方法 —— 投影寻踪法中，存在着大量的难于解决的求分布问题. 用 Bootstrap 逼近提供了一种在实用上可行的处理方法，而这种逼近又带来了一些需要从理论的角度研究的问题，因此，在这个意义下也可以说，Bootstrap 方法（以及由它引伸而来的随机加权法）与高维数据统计分析有着密切的关系.

二　投影寻踪法
(Projection Pursuit Method, PP 方法)

投影寻踪法，简称 PP 法，是一种比较典型的针对数据维数很高而设计的方法. 这个方法可溯源到本世纪 70 年代初期. 1974 年，Friedman 和 Tukey[13] 将这方法用于探索性数据分析，并创立了“Projection Pursuit”这个称呼. 1985 年，Huber 发表了一篇关于 PP 方法的综述性文章[1]，对到当时为止的发展情况作了总结并提出了若干有待解决的问题. 在中文文献方面，成平和李国英等在 1986 年有综述文章[2] 和讲义[3]，对这个领域作了比较细致的介绍，可以参考.

从一个角度看，PP 方法的基本思想，在人们习知的统计方法

中，早已有所反映．例如多元分析中的主成分分析方法．设 $X_1,\cdots,X_n$ 是 p 维随机变量 X 的 n 个观察值．对一个 p 维单位向量 a，计算 $Y_i=a^{\tau}X_i,\ i=1,\cdots,n$（$a$ 为列向量，a^{τ} 指其转置）．算出 $Y_1,\cdots,Y_n$ 的样本方差 $S_1^2(a)$，它与 a 有关．找 $a=a_1$，使 $S_1^2(a_1)=\sup\limits_{\|a\|=1}S_1^2(a)$，则 $a_1^{\tau}x$ 是第一即最大的主成分而 a_1 为相应的投影方向．$S_1^2(a)$ 在 $a=a_1$ 处最大，表示将样本 $X_1,\cdots,X_n$ 往 a_1 这个方向投影．其结果包含了最多的信息，a_1 可称为最优或最有兴趣的投影方向．在定下 a_1 以后，就可以继续定出第二、第三、… 有兴趣的投影方向，相应于第二、第三、… 主成分．在有些情况下，尽管原来的数据维数 p 很高，但很少几个主成分即能包含原数据信息中的主要部分．这使我们可以在一个比 p 低得多的维数的空间中去考虑问题，而实质上并无损失．从这一点看，主成分法无非是一种降维法．

以上的概念推广一步，即得到 PP 统计量这个基本概念．即在向 a 方向投影而得到 $a^{\tau}X_1,\cdots,a^{\tau}X_n$ 后，我们不一定去计算其样本方差，而是选择一个适当的 n 元函数 Q，去计算 $Q(a^{\tau}X_1,\cdots,a^{\tau}X_n)$．这样的函数 Q 可称为"投影指标"(Projection Index)，其选择是基于尽量有利于发现数据的基本结构及其有兴趣的特征这一点．一旦定了 Q，就可以寻找 a 之值 a_*，使

$$Q(a_*^{\tau}X_1\cdots,a_*^{\tau}X_n)=\sup_{\|a\|=1}Q(a^{\tau}X_1,\cdots,a^{\tau}X_n),$$

这种最大值常称为 PP 统计量．

PP 方法目前已被用于许多重要的统计问题中，在国内也进行了多方面的研究，取得了不少结果．此处因篇幅所限，主要只就假设检验和尾部概率的阶的估计问题作较仔细的介绍．更进一步的细节请参阅所引的参考文献．

1. PP 检验

Huber 在[1]中曾指出，PP 提供了迄今为止最有效的方法以将一些一维情况下的做法移植于多维．检验位置参数 m 是一个简

单例子. 为要检验 $m = m_0$，在一维情况下有 t 统计量

$$t_n = t_n(X_1,\cdots,X_n) = (\overline{X}_n - m_0)/S_n,$$

这里$\overline{X}_n$ 和 S_n 分别是样本均值和样本方差. 如果 m 是多维的，则注意到

$$m = m_0 \Leftrightarrow a^\tau m = a^\tau m \text{，对任何 } \|a\| = 1 \text{ 的 } a.$$

因此，统计量

$$\begin{aligned} T_n &= \sup_a \left\{ \left| a^\tau \overline{X}_n - a^\tau m_o \right| / \sqrt{a^\tau Z_n a} \right\} \\ &= \sup_a \left| t_n(a^\tau X_1,\cdots,a^\tau X_n) \right| \end{aligned}$$

提供了一个合宜的检验法. 使用经验过程的理论（见[4]），可以在原假设 $H_o: m = m_0$ 之下导出 T_n 的渐近分布. 类似本例的处理方法把一批一维方法推广到多维. 人们称上文定义的 T_n 为 t_n 的 PP 统计量. 类似的情况也使用这个称呼.

我国统计工作者对两个多维位置参数向量相等的检验问题，提出了一系列的 PP 检验统计量（见[5]—[8]），其中有的基于 Mann-Whitney 检验，稳健的 t 型检验，及以 L 和 R 统计量为投影指标的PP检验统计量. 在这些工作中也讨论了散布矩阵的检验问题. 麻烦的是，这些统计量在原假设下的渐近分布依赖于总体的确切分布. 为此，在[8]中讨论了用“Bootstrap”方法（见本文第四节）去逼近 PPL- 统计量的问题，并提供了若干模拟结果. 在工作[9]中则讨论了 PPU- 统计量的问题. 对这种类型的统计量，用通常的经验过程方法难于导出渐近分布，为此引入了“U- 经验过程”的概念并导出其渐近分布，然后藉助于它去推导 PPU- 统计量的渐近分布. 使用这个方法，得出了一些一、两样本位置参数向量，散布矩阵及多维空间中独立性的检验方法.

下面来考虑一下多维拟合优度检验问题，设 $X_1,\cdots,X_n$ 为抽自 d 维分布 F 的独立同分布样本. 以 F_n 记 $X_1,\cdots,X_n$ 的经验分布函数，以 P 和 P_n 分别记相应于 F 和 F_n 的概率测度，以 F^a 记 $a^\tau X$ 的

分布函数，相应的概率测度记为 P^a，类似地有 F_n^a 和 P_n^a. 在工作[10] 中提出了 PPχ^2 统计量：

$$Z_n = \sup_{\|a\|=1} \sum_{i=1}^{m} n[P_n^a(S_i) - P^a(S_i)]^2 / P^a(S_i),$$

这里 $\{S_1, \cdots, S_m\}$ 是 R' 的一个分割. 在原假设成立之下求出了 Z_n 的渐近分布. 另外，在工作[11] 和[12] 中则分别引进了 PP Cramer-Von Mises-Smirnov 统计量和 PP Neyman 统计量：

$$W_n = \sup_{\|a\|=1} \int_{-\infty}^{\infty} (F_n^a(t) - F^a(t)^2 \mathrm{d}F^a(t),$$

$$\widetilde{W} = \int_{\|a\|=1} \int_{-\infty}^{\infty} (F_n^a(t) - F^a(t))^2 \mathrm{d}F^a(t) \mathrm{d}H(a),$$

$$K_n = \sup_{\|a\|=1} \sum_{j=1}^{m} \Big[\sum_{i=1}^{n} \pi_j(F^a(a^\tau X_i)) \Big]^2 / n,$$

这里 $\pi_0 = 0, \pi_1, \pi_2, \cdots$ 是区间(0,1) 上的正交多项式，$H(a)$ 表示 n 维单位超球面上的均匀分布. 在上述工作中分别导出有关统计量的渐近分布，考虑了两种类型的 Bootstrap 逼近，证明了其相合性：这些 Bootstrap 统计量与原来的 PP 统计量有同一的渐近分布. 这里提出的 PP Neyman 统计量比 Friedman 在 1987 年[13] 中提出的一种统计量要广泛一些，虽然二者略有不同. 以上引述的工作只研究了有关统计量在原假设下的分布. 若要考虑相应检验的功效，则必须考虑对立假设下的渐近分布问题. 工作[14] 讨论了这个问题，有关结果不在此引述了.

2. PP 统计量的尾部概率

许多PP统计量在原假设下的渐近分布，与总体的确切分布有关，当后者未知(例如在检验多维分布 F 的期望 $m = m_0$ 时，知道 m 不等于知道 F) 时，仍无法定出基于 PP 统计量的大样本分布的检验临界值. 为解决这个困难，一个做法是用 Bootstrap，另一个是对 PP 统计量的尾部概率作出理论上的估计.

统计量的尾部概率的估计问题是概率论中一个古老而基本的

问题.其所用技巧,对处理 PP 统计量尾部概率问题当然有借鉴作用,但也涉及一些新工具和新技巧.近几年来这个问题吸引了一些学者的注意,我国统计工作者在这方面做了一些工作.有关其他国家的工作可参看[15]—[21].下面介绍一下我国在这方面的若干重要工作.

a. 以 $\mathscr{F}$ 记 R^d 上的一个函数类,其图形所成的类有 ν 阶多项式分辨率(见[4]第 2 章).Alexander[15]证明了当 $\lambda \geqslant 8$ 时有

$$P(\sup_{F\in\mathscr{F}}|P_nf-Pf|>\lambda)\leqslant 16\lambda^{2^{12}\nu}\exp(-2\lambda^2),$$

这里 P,P_n 及下文的 F^a,F_n^a 等在前面已解释过.又对 R^d 上一个概率测度 Q,Qf 的意义是 $\int_{R^d} f\mathrm{d}Q$.

工作[24]改进了这个结果.例如,证明了:对任何 $\varepsilon\in(0,1/4)$,当 $\lambda\geqslant\lambda_0=\lambda(\nu,\varepsilon)$ 时,上述不等式中的因子 $16\lambda^{2^{12}\nu}$ 可以 $\lambda^{4(\nu-1)(1+\varepsilon)/(1-2\varepsilon)}$ 去代替.[24]中还针对某些特殊情况给出了更为精确的估计.

b. 对 PP K-S(Kolmogorov-Smirnov)统计量,Ohrvik 在其工作[18]和[19]中,对正态和其他椭球等高分布进行了不少模拟工作,而对尾部概率

$$\xi(n,d,\lambda)\equiv P(\sqrt{n}\sup_{a,t}|F_n^a(t)-F^a(t)|>\lambda)$$

得出了两个经验不等式.

Huber[20]从理论上研究了这个问题,对总体分布球对称的情况证明了

$$\xi(n,d,\lambda)\leqslant 2(en/d)^d\exp\{-2(\lambda-d/\sqrt{n})^2\},\quad \lambda>d/n. \tag{1.1}$$

基于Ohrvik 的模拟结果,Huber 提出了如下的猜测:

$$\xi(n,d,\lambda)\leqslant C(d)\exp(-2\lambda^2), \tag{1.2}$$

此处 $C(d)$ 只依赖于 d,而不依赖 λ 和 n.

我国统计工作者在一系列的工作中研究了这个问题(见

[23]—[29]). 他们获得的一个重要结果是:对某些类型的 d 维分布当 $n \geqslant n_0(\lambda)$ 时有

$$C_1(d)\lambda^{2(d-1)}\exp(-2\lambda^2) \leqslant \xi(n,d,\lambda) \leqslant C_2(d)\lambda^{2(d-1)}\exp(-2\lambda^2). \tag{1.3}$$

由此式可知,Huber 的猜测(1.2)是错误的.(1.3)的右边改进了 Huber 的工作(1.1),它不仅对椭球对称分布成立,也对某些对称稳定分布成立.(1.3)的左边甚至对更广的分布类也成立.

c. 在工作[23]和[29]中研究了 m 维投影PP K-S统计量的尾部概率.记

$$\mathscr{F}_m = \{I(Ax \leqslant t): A \text{ 为 } m \times d \text{ 矩阵}, t \in R^m\},$$

$$\xi_m(n,d,\lambda) = P\{\sup_{f \in \mathscr{F}_m} |\sqrt{n}(P_n f - Pf)| > \lambda\}.$$

证明了:若 P 为椭球对称,$P(X=0)=0$,则对任何 $\lambda \geqslant 2$ 存在 $n(\lambda)$,使当 $n \geqslant n(\lambda)$ 时,有

$$\xi_m(n,d,\lambda) \leqslant C\lambda^{2(dm-1)}\exp(-2\lambda^2).$$

在这些工作中,也讨论了对更一般的 $P\xi_m(n,d,\lambda)$ 的上界和下界问题.

d. 在工作[30]中把 Von-Mises 统计量和 PP 概念结合起来,引进了两个拟合优度检验统计量:

$$V_n = \int_{\|a\|=1}\int_{-\infty}^{\infty}[\sqrt{n}(F_n^a(t) - F^a(t)]^2 \mathrm{d}F^a(t)\mathrm{d}H(a),$$

H 为 R^d 的单位球面上的均匀分布,以及

$$W_n = \sup_{\|a\|=1}\int_{-\infty}^{\infty}[\sqrt{n}(F_n^a(t) - F^a(t)]^2 \mathrm{d}F^a(t).$$

在[30]中证明了:在某些较弱的条件下,当 $\lambda \geqslant 1$ 时存在 $n(\lambda)$,使当 $n \geqslant n(\lambda)$ 时有

$$P(V_n > \lambda) \leqslant C\lambda^{-1/2}[\ln\lambda]^+ e^{-\lambda\pi^2/2},$$

而当 W 为椭球对称时有

$$P(W_n > \lambda) \leqslant C(d)\lambda^{2d-1/2}e^{-\lambda\pi^2/2}.$$

e. 我们考虑了加权的PP型Cramér-von Mises统计量的极限分布和尾部概率的渐近表达式.令

$$\widetilde{V}_n = n\int_{\|a\|=1}\int (F_n^a(t) - F^a(t))^2\lambda(F^a)\mathrm{d}F^a(t)\mathrm{d}H(a)$$

在 F 服从 $\|a\| = 1$ 上均匀分布 $H(a)$, $\lambda(F^a(t)) = (1-t^2)^{-(m-2)}$ 时我们找到了 $\widetilde{V}_n$ 极限分布的明确表达式.即与

$$\sum_{j=1}^{\infty} \lambda_j \chi_j^2(h_j(1))$$

有同一分布.

$$\text{此处}\quad \lambda_j = \left[\frac{\Gamma\left(\frac{m}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{m-1}{2}\right)}\right]^2 \cdot \frac{1}{j(j+m-2)} \cdot \frac{1}{h_j(1)},$$

$$h_j(1) = \binom{m-2+j}{m-2} + \binom{m-3+j}{m-2},$$

$x_j^2(h_j(1))$ 为自由度为 $h_j(1)$ 的 χ^2 随机变量,而且相互独立.

在 F 服从正态分布 $N(\mu,\Sigma)$, $\lambda(F^a) = \varphi^{-2}(t)\left(\varphi(t) = \frac{1}{\sqrt{2\Pi}}e^{-\frac{t^2}{2}}\right)$.令

$$Q(x,y) = \int_{-\infty}^{\infty} [\,\mathrm{I}_{(x<t)} - \Phi(t)][\,\mathrm{I}_{(y<t)} - \Phi(t)]\varphi^{-1}(t)\mathrm{d}t.$$

在 μ,Σ 已知或未知而代以样本均值和样本协方差时,我们都得到 $\widetilde{V}_n - \int_{\|a\|=1} \frac{1}{n}\sum_{i=1}^{n} Q(a^\tau x_i,\ a^\tau x_i)\mathrm{d}H(a)$ 极限分布尾部展开式,也是独立 χ^2 变量的线性组合的分布.从而可用来检验总体是否服从正态分布或某一指定的正态分布.

上述结果可参见中国科学院系统科学研究所崔恒建1993年博士论文.

三 数论方法在数理统计问题中的应用

这方面的研究在我国也是近十年来兴起的. 它以我国著名数学家华罗庚教授和王元教授提出的方法(见[31])为基础. 在某些高维数据分析问题中,这种方法具有节省计算量的突出优点,具有较好的精度. 目前这方面的研究成果主要有二:一是涉及有关高维随机向量的分布的数值计算问题,一是在涉及众多因子和水平的情况下的试验的设计问题. 研究工作的主要成果,总结在[32]和[33]两篇文献中. 此处举例对其要点作一些介绍.

华-王关于将数论方法用于数值积分的方法,是基于"一致分布点列"这个概念. 设有一个定义在 $I=[0,1]^s$ 上的 s 变元的足够光滑的函数 f,要计算积分 $\int_I f(X)\mathrm{d}x$. 设 $P_n=\{X_1^{(n)},X_2^{(n)},\cdots,X_{k_n}^{(n)}\}$, $n=1,2,\cdots$,是 I 中的一串点列,满足 $k_n\to\infty$ 当 $n\to\infty$. 在 I 内任意指定一点 $b=(b_1,\cdots,b_s)$,并且对于任何 $c=(c_1,\cdots,c_s)$,约定 $c\leqslant b$ 是指 $c_i\leqslant b_i, i=1,\cdots,s$. 以 $N(n,b)$ 记 P_n 中落在区间 $[0,b]$ 上的点的个数,即满足条件 $x_i^{(n)}\leqslant b$ 的 i 的个数, $i=1,\cdots,k_n$,而计算

$$D_n=\sup\{|N(n,b)/k_n-b_1\cdots b_s|:b\in I\},$$

如果

$$\lim_{n\to\infty}D_n=0,$$

则称 $\{P_n\}$ 是 I 上的一个一致分布点列. 华-王的工作证明了:若 $\{P_n\}$ 是一致分布点列而 f 足够光滑,则 $\sum_{i=1}^{k_n}f(x_i^{(n)})/k_n$ 是积分 $\int_I f\mathrm{d}x$ 的一个精度好而计算量省的近似.

这种结果立即可用于有关的应用统计问题,其中涉及到有关的量的分布计算. 但在现实问题中,积分区域常不是一个长方体,

因而存在着理论和技术上的困难. 初一看，若积分区域 A 为有界，似可用如下的方法去处理：作一个正立方体 $I \supset A$，定义

$$g(x) = \begin{cases} f(x), \text{当 } x \in A; \\ 0, \text{当 } x \in I - A. \end{cases}$$

然后去计算 $\int_I g(x)\mathrm{d}x = \int_A f(x)\mathrm{d}x$. 这个做法的问题在于如上所定义的函数 g 在 A 的边界处一般不连续，这将对精度产生较大的影响. 更进一步，在统计问题中有时需要计算定义于 R^s 的函数 f 在 R^s 上的某一曲面上的积分，这时更无可能采用上面的做法了.

好在，在应用统计问题中，所涉及的积分区域往往有比较特殊的形状，这使我们有可能通过适当的变换，把积分区域 A 一一且光滑地变换成 R^s 中的或更低维的 $R^t(t < s)$ 中的单位立方体，变换后的被积函数仍足够光滑. 通过这个方式就可以把一般区域上的数值积分化为正立方体上数值积分的情况. 从另一种角度看，这可以看成是在一个一般区域或曲面（当然其形状要相当"规则"）上寻找一致分布点列的方法. 在工作[32]，[33] 中详细描述和论证了这个过程，并将其应用于一些在实用上有重要性的特例. 以下考察其中的几个例子作为说明.

先看一个合金钢冶炼的例子. 设 x 为一个 s 维列向量，其各分量表示合金钢中所含某些化学成分的量，而 $\mu(x)$ 则表示相应的合金钢的质量. 假定 x 和 $\mu(x)$ 的关系由一个线性回归模型来表达：

$$E[\mu(x)] = \hat{\mu}(x) = a + Bx,$$

这里 a 为 t 维列向量，B 是 $t \times s$ 阶的回归系数矩阵. 设对每个 $x \in K = [a_1, b_1] \times \cdots \times [a_s, b_s]$，$\mu(x)$ 有 t 维正态分布 $N_t(a + Bx, \Sigma)$，Σ 为 t 阶正定方阵. 在实际问题中，a，B 和 Σ 一般都是需要估计未知参数. 我们假定，已通过样本并使用最小二乘或其他方法得出了其估计，并为方便计，将估计值仍记为 a，B 和 Σ.

设 $T_i, i = 1, \cdots, t$，是一些指定的数值. 当 $\mu(x)$ 的各分量 $\mu_1, \cdots, \mu_t$ 依次大于 $T_1, \cdots, T_t (\mu_i > T_i, i = 1, \cdots, t)$ 时，该合金钢的质

量是合格的. 则"该合金钢的质量合格"这个事件的概率是

$$p(x) = \int_{T_1}^{\infty} \cdots \int_{T_t}^{\infty} n_t(y_1, \hat{\mu}(x), \Sigma)\,dy,$$

这里 $n_t(y, \hat{\mu}(x)\Sigma$ 是 t 维正态分布 $N_t(\hat{\mu}(x), \Sigma)$ 的概率密度函数.

为计算这个概率,可取 $A_i > T_i$ 足够大,$i = 1, \cdots, t$(其具体值可依对计算精度的要求去估计之),而将上述积分用一个有界域上的积分去近似地逼近之:

$$p(x) \approx \int_{T_1}^{A_1} \cdots \int_{T_t}^{A_t} n_t(y, \hat{\mu}(x), \Sigma)\,dy.$$

于是,若$\{b_k = (b_{k1}, \cdots, b_{kt})\}$是单位正方体$[0,1]^t$上的一致分布点列,则$\{Z_k = (Z_{k1}, \cdots, Z_{kt})\}$,其中

$$Z_{ki} = T_i + (A_i - T_i)b_{ki}, i = 1, \cdots, t$$

是$[T_1, A_1] \times \cdots \times [T_t, A_t]$上的一致分布点列,而 $p(x)$ 可以用公式

$$p(x) \approx \prod_{i=1}^{t} (A_i - T_i) \frac{1}{n} \sum_{k=1}^{n} n_t(Z_k, \hat{\mu}(x), \Sigma)$$

去近似地计算.

再看一个计算一种特殊的多维分布的混合矩的例子.

设有 N 维随机向量$(X_1, \cdots, X_N)^r$,满足条件 $X_i \geqslant 0, i = 1, \cdots, N$ 而 $X_1 + \cdots + X_N = 1$. 换言之,$X_1, \cdots, X_N$ 可视为一物质中所含 N 种成分的比率,记 $X_{-N} = (X_1, \cdots, X_{N-1})^r$,作变换

$$Y = \log(X_{-N}^r / X_N) = \left(\log \frac{X_1}{X_N}, \cdots, \log \frac{X_{N-1}}{X_N}\right)^r,$$

该变换将单纯形

$$T_N = \{(t_1, \cdots, t_N) : t_1 > 0, \cdots, t_N > 0, t_1 + \cdots + t_N = 1\}$$

一对一地变换到 R^{N-1}. 称$(X_1, \cdots, X_N)$服从可加罗吉斯蒂(additive logistic)正态分布 $AN_{N-1}(\mu, \Sigma)$,如果经变换后的随机向量 Y 服从正态分布 $N_n(\mu, \Sigma)$(参见[32]).

由上述变换,利用行列式公式

$$\begin{vmatrix} \frac{1}{X_N}+\frac{1}{X_1} & \frac{1}{X_N} & \frac{1}{X_N} & \cdots & \frac{1}{X_N} \\ \frac{1}{X_N} & \frac{1}{X_N}+\frac{1}{X_2} & \frac{1}{X_N} & \cdots & \frac{1}{X_N} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \frac{1}{X_N} & \frac{1}{X_N} & \frac{1}{X_N} & \cdots & \frac{1}{X_N}+\frac{1}{X_{N-1}} \end{vmatrix} = \prod_{i=1}^{N} x_i^{-1}$$

$$\left(\sum_{i=1}^{N} X_i = 1\right)$$

易得 $AN_{N-1}(\mu,\Sigma)$ 的概率密度函数为($|\Sigma|$ 为 Σ 的行列式)

$$(2\pi)^{-(N-1)/2}|\Sigma|^{-1/2}\prod_{i=1}^{N} X_i^{-1}\exp\left\{-\frac{1}{2}\left(\log\frac{x_{-N}}{X_n}-\mu\right)^{\tau}\right.$$

$$\left.\cdot\,\Sigma^{-1}\left(\log\frac{x_{-N}}{x_N}-\mu\right)\right\},$$

从而得到 X 的 $(t_1,\cdots,t_N)$ 阶混合矩,即

$$E(X_1^{t1}\cdots X_N^{tN}) = (2\pi)^{-(N-1)/2}|\Sigma|^{-1/2}$$

$$\cdot\int_{T_N}\prod_{i=1}^{N} x_i^{t_i-1}\exp\left\{-\frac{1}{2}\left(\log\frac{x_{-N}}{x_N}-\mu\right)^{\tau}\Sigma^{-1}\left(\log\frac{x_{-N}}{x_N}-\mu\right)\right\}\mathrm{d}V,$$

这里 $\mathrm{d}V$ 为 T_n 的体积元. 更确切地,它是 R^{N-1} 的子集合

$$\left\{(x_1,\cdots,x_{N-1}):x_i>0,i=1,\cdots,N-1;\sum_{i=1}^{N-1}x_i<1\right\}$$

的体积元.

使用文献[32]中例4的变换,可以将上述积分转化为 R^{N-1} 中的单位正立方体 I_{N-1} 上的积分,此处我们不仔细写出变换的仔细计算过程,仅将结果引述如下:

$$E(X_1^{t_1}\cdots X_N^{t_N}) = C\int_{I_{N-1}}\prod_{i=1}^{N-1}(C_i^{2t_i-1}S_i^{2(t_{i+1}+\cdots+t_N)-3})Q(\varphi)\mathrm{d}\varphi,$$

其中$\mathrm{d}\varphi=\mathrm{d}\varphi_1\cdots\mathrm{d}\varphi_{N-1}$, $C=(2/\pi)^{-(N-1)/2}|\Sigma|^{-1/2}\sqrt{N}$,以及

$$S_i=\sin(\pi\varphi_i/2),C_i=\cos(\pi\varphi_i/2),i=1,\cdots,N-1,$$

$$Q(\varphi)=\exp\left\{-\frac{1}{2}(g(\varphi)-\mu)^{\tau}\sum{}^{-1}(g(\varphi)-\mu)\right\},$$

而

$$g(\varphi) = 2\Big(\log C_1 - \sum_{i=1}^{N-1}\log S_i, \log C_2 - \sum_{i=2}^{N-1}\log S_i, \cdots, \log C_{N-1} - \sum_{i=N-1}^{N-1}\log S_i\Big).$$

然后可用前面交代过的方法，通过单位正立方体中的一致分布点列去近似计算上述积分.

下面我们简单地介绍一下一致分布点列在试验设计中的应用. 这种设计的基本思想，是把试验点列尽可能均匀地配置在试验区域内. 如果有 n 个连续取值的因子 $A_1, \cdots, A_n$，分别可独立地在有界区间$[a_1, b_1], \cdots, [a_n, b_n]$上取值，则均匀设计在于要求在$[a_1, b_1] \times \cdots \times [a_n, b_n]$上定出一个一致分布的点列，而这可以转化为单位正方体 I_n 的情形. 对混料试验，其试验区域为$\{(x_1, \cdots, x_n): x_i \geqslant 0, i = 1, \cdots, n, x_1 + \cdots + x_n = 1\}$，可以通过正立方体上的一致分布点列，作出上述单纯形上的一致分布点列，详见[31]和[32]. 均匀设计的思想也可用于因子只取有限个水平的情况. 当有 s 个因子，每因子有 n 水平时，全面试验需做 n^s 次，即使使用某种部分实施，例如正交拉丁方，所需试验次数往往仍嫌过大. 这时，均匀设计在某些问题中不失为一个可行的替代者.

均匀设计 $U_n(b^c)$ 表示这样一个设计：它需做 n 次试验，每因子有 b 水平，而能安排的因子的最大个数为 c. 例如 $U_{11}(11^6)$ 的表为

行号 \ 列号	1	2	3	4	5	6
1	1	2	3	5	7	10
2	2	4	6	10	3	9
3	3	6	9	4	10	8
4	4	8	1	9	6	7
5	5	10	4	3	2	6
6	6	1	7	8	9	5
7	7	3	10	2	5	4

续上表

行号＼列号	1	2	3	4	5	6
8	8	5	2	7	1	3
9	9	7	5	1	8	2
10	10	9	8	6	4	1
11	11	11	11	11	11	11

这个表可以排至多6个11水平因子作11次试验.如因子数$h<6$,则需要在表中适当选择h列,为此目的附有一张表,指导如何选择列.在本例中是:

$h=2:(1,5)$; $h=3:(1,4,5)$; $h=4:(1,2,4,5)$; $h=5:(1,2,3,4,5)$.

例如,有三个各有11水平的因子A,B和C,按上述规定,把A、B、C分别排在$U_{11}(11^6)$表的第1、4、5列,得到试验的11个处理组合:

$A_1B_5C_7$, $A_2B_{10}C_3$, $A_3B_4C_{10}$, $A_4B_9C_6$, $A_5B_3C_2$, $A_6B_8C_9$, $A_7B_2C_5$, $A_8B_7C_1$, $A_9B_1C_8$, $A_{10}B_6C_4$, $A_{11}B_{11}C_{11}$.

均匀设计表的制作方法,在文献[33]中有仔细描述,读者可以参看.数论方法还成功地用于估计几何概率及最优化等,此处都不细述了.

四　随机加权法

近年来,再抽样法引起统计学者的注意.一般地说,所谓再抽样法,就是把手头已有的样本作为总体,从中进行抽样.初期的想法是这样可以从一组样本对同一参数产生许多估计,以有助于估计(某统计量的)方差或进行区间估计等.古典的Jacknife是一个

重要的例子.

1979 年,Efron 发表了一篇后来被证明是影响很大的文章[34],其中提出了"Bootstrap 方法"的概念:

有了样本 $X_1,\cdots,X_n$ 后,考虑其经验分布 F_n,然后在 F_n 中进行抽样.这方法的主要目的在于近似地求某些量的分布,其精确分布很难或无法求得.一般程式如下:设 $X_1,\cdots,X_n$ 为抽自分布 F 中的独立同分布样本,记 $X_{(n)}=(X_1,\cdots,X_n)$.要求某个量 $Q(X_{(n)},F)$ 的分布,Q 为 $X_{(n)}$ 和 F 的某个已知泛函.设 $X_{(n)}^*=(X_1^*,\cdots,X_n^*)$ 为自 $X_{(n)}$ 的经验分布 F_n 中抽出的独立同分布样本,即以 $Q(X_{(n)}^*,F_n)$ 的分布作为 $Q(X_{(n)},F)$ 的分布的近似.这就是 Efron 提出的 Bootstrap 方法的要旨.值得注意的是,因分布 F_n 已知,原则上我们可随心所欲地抽出任意多个 $X_{(n)}^*$ 的样本,因而在经验上可以把 $Q(X_{(n)}^*,F_n)$ 的分布决定到任意指定精度的范围内.

举一个例来说,总体分布 F 有期望 θ,$X_1,\cdots,X_n$ 为从 F 中抽出的独立同分布样本,以 $\overline{X}=\sum_{i=1}^{n}X_i/n$ 作为 θ 的估计.为找 θ 的基于 $\overline{X}$ 的区间估计,需要知道 $\overline{X}-\theta$ 的分布,当对总体分布 F 并无特殊假定时,这无法做到.可是根据上述 Bootstrap 方法的原则,可以用 $\overline{X}^*-\overline{X}$ 的分布近似地作为 $\overline{X}-\theta$ 的分布,其中 $\overline{X}^*=\sum_{i=1}^{n}X_i^*/n$. Bootstrap 理论的一个基本问题是证明当样本大小 $n\to\infty$ 时,$Q(X_{(n)}^*,F_n)$ 的分布确实无限地接近 $Q(X_{(n)},F)$ 的分布.这方面已有了很多结果,不能一一引述.拿上例而言,如果总体分布 F 的方差 σ^2 有限,则 $\sqrt{n}(\overline{X}^*-\overline{X})$ 的分布和 $\sqrt{n}(\overline{X}-\theta)$ 的分布当 $n\to\infty$ 时都以正态 $N(0,\sigma^2)$ 为极限,因而二者无限接近.

这种结果可立即用于求 θ 的区间估计.设经过大量抽样已相当精确地定出了 $\overline{X}^*-\overline{X}$ 的分布 G.找 G 的上、下 $\alpha/2$ 分位点 u 和 v,即 $G(u)=1-\alpha/2$,$G(v)=\alpha/2$.则近似地有 $P(v\leqslant\overline{X}-\theta\leqslant u)$

$\approx 1-\alpha$ 而得到 θ 的置信系数 $1-\alpha$(近似)的区间估计为 $[\overline{X}-\overline{u}, \overline{X}-v]$.

但这种做法有一个问题,经验证明:用 Bootstrap 方法作出的区间估计,其真实的置信系数往往比预先指定的为小,这个现象严格的理论解释尚不很清楚.一种粗略的看法如下:$X_1,\cdots,X_n$ 的经验分布的方差为 $\sum_{i=1}^{n}(X_i-\overline{X})^2/n$. 其期望值为 $\frac{n-1}{n}\sigma^2$,偏低于总体方差 σ^2. 样本大小 n 愈小,偏低愈显著.故拿均值 θ 的估计这个例子说,可以预期,一般讲 $\overline{X}^*-\overline{X}$ 的上 $\alpha/2$ 分位数将倾向小于 $\overline{X}-\theta$ 的上 $\alpha/2$ 分位数,下 $\alpha/2$ 分位数(为负)则相反.这样,从 Bootstrap 分布所算出的置信区间,将包含在基於 $\overline{X}-\theta$ 的精确分布所算出的置信区间之内,因而也就导致置信系数的降低.

为了对付这个问题,郑忠国在一系列工作[35]—[38]中提出和发展了一种以 Bootstrap 方法的思想为基础的方法 —— 随机加权法(Rubin 的工作[39]中也包含了这种思想,但未加以发展).要讲清这方法的思想并明了其与 Bootstrap 思想的联系,我们考察一个例子 ——Bootstrap 均值

$$\overline{X}^* = (X_1^* + \cdots + X_n^*)/n.$$

按定义,X_i^* 可表为

$$X_i^* = \eta_{i1}X_1 + \eta_{i2}X_2 + \cdots + \eta_{in}X_n,\ i = 1,\cdots,n,$$

其中 $\{\eta_{ij}\}$ 全体是与 $\{X_i\}$ 独立,$(\eta_{i1},\cdots,\eta_{in}),i=1,\cdots,n$ 独立同分布,其公共分布为

$$P\{\eta_{i1},\cdots,\eta_{in}) = (0,\cdots,0,1,0,\cdots,0)\} = 1/n,\ k = 1,\cdots,n,$$

其中 1 处在第 k 个位置,这样,若记

$$\eta_j = \sum_{i=1}^{n}\eta_{ij}/n,\ j = 1,\cdots,n,$$

则

$$\overline{X}^* = \eta_1X_1 + \eta_2X_2 + \cdots + \eta_nX_n.$$

所以,Bootstrap 均值$\overline{X}^*$ 其实是 $X_1,\cdots,X_n$ 的一种加权和,权是随机的,但是一种特殊的权:$n\eta_1,\cdots,n\eta_n$ 必须为非负整数,其和为 n.

郑忠国把上述形式加以拓广,即在上述$\overline{X}^*$ 的表达式中,$(\eta_1,\cdots,\eta_n)$ 可以是某些其分布经过适当选择的随机向量,它一般是连续型的,即权 η_i 不必局限于 m/n 的形状,而可取任何非负值,甚至 η_i 的和也可以不必为1(仍假定$(\eta_1,\cdots,\eta_n)$与$(X_1,\cdots,X_n)$独立). 这样一来,Bootstrap 分布成为连续型的,较易处理,且更重要的是,由于$(\eta_1,\cdots,\eta_n)$ 的分布有选择余地,可根据不同问题的需要进行挑选而达到较好的效果. 在实施方面,当然要考虑到$(\eta_1,\cdots,\eta_n)$ 的样本要便于通过随机模拟产生. 迄今在这方面的工作中最常用到的两种选择是:

1. $(\eta_1,\cdots,\eta_n)$ 服从 Dirichlet 分布 $D(4,4,\cdots,4)$,即:$\eta_1+\cdots+\eta_n=1$,而$(\eta_1,\cdots,\eta_{n-1})$ 的密度在集合

$$\left\{(u_1,\cdots,u_{n-1}):u_1\geqslant 0,\cdots,u_{n-1}\geqslant 0,\sum_{i=1}^{n-1}u_i\leqslant 1\right\}$$

上为$(\Gamma(4n)/\Gamma^n(4))u_1^3\cdots u_{n-1}^3(1-u_1-\cdots-u_{n-1})^3$,在其外为 0.

2. $\eta_1,\cdots,\eta_n$ 独立同分布,有公共分布 $\Gamma(4)$,即有密度函数 $\frac{1}{6}x^3e^{-x},x>0$. 在此权和 $\eta_1+\cdots+\eta_n$ 当然不必为 1.

对一些分布的模拟结果显示:用这种随机加权法所造出的区间估计,与用原始的 Bootstrap 方法造出的区间估计相比,其置信系数更接近于预先指定的值. 兹引述[35] 中的一例如下:总体分为 $N(0,1)$,故被估计的期望的真值为 0. 样本大小 $n=200$. 置信系数指定为 0.90. 重复 500 次. 对原始的 Bootstrap 方法来说,所作出的置信区间覆盖 0 的频率为 0.814. 用加权平均法,则按上述第一种方式取$(\eta_1,\cdots,\eta_n)$的分布(即$(\eta_1,\cdots,\eta_n)\sim D(4,4,\cdots,4)$)时,这个数字为 0.878,如用第二种方法取$(\eta_1,\cdots,\eta_n)$ 的分布,这个数字为 0.870. 这两个数字都很接近预定值 0.90,且都比原始

Bootstrap 法得出的 0.814 要好得多.

除均值外,郑忠国也把这个方法使用于另一些更复杂的、在应用上重要的情况,如 t 统计量,线性回归模型,U 统计量等.样本也不一定限制为独立同分布的.理论上的讨论限于大样本性质方面,即当样本较大,趋于无穷时,随机加统计量的极限分布和渐近展开.我们引述几个重要的结果如下:

1. 均值的随机加权

设 $X_1,\cdots,X_n$ 为抽自分布 F 的独立同分布样本,作随机加权平均值

$$E_n = \sum_{i=1}^{n}(X_i - \overline{X})\eta_i,$$

这里 $(\eta_1,\cdots,\eta_n) \sim D(4,\cdots,4)$. 给定 $X_1,\cdots,X_n$ 的条件下,E_n 的条件期望为 0,条件方差 $(\sigma_n^*)^2$ 为

$$(\sigma_n^*)^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2/[n(4n+1)].$$

以 σ^2 记 F 的方差,设 $0 < \sigma^2 < \infty$,并设 $\int_{-\infty}^{\infty}|x|^3 \mathrm{d}F(x) < \infty$,则以概率 1 成立

$$\sup_y \left| P\left(\frac{\sqrt{n}\,(\overline{X}-\theta)}{\sigma} \leqslant y\right) - P^*\left(\frac{E_n}{\sigma_n^*} \leqslant y\right) \right| = o\left(\frac{1}{\sqrt{n}}\right),$$

此处 θ 为分布 F 的期望,P^* 表示在给定 $X_1,\cdots,X_n$ 之下计算条件概率.如前所述,这结果可用构造 θ 的置信区间.值得注意的是上述逼近的阶为 $o\left(\frac{1}{\sqrt{n}}\right)$. 如采取 $D(1,\cdots,1)$ 作为 $(\eta_1,\cdots,\eta_n)$ 的分布,则逼近的阶由 $o\left(\frac{1}{\sqrt{n}}\right)$ 降为 $o(1)$. 这里看出权的选择的影响及适当地选择权带来的好处.

2. t 统计量的随机加权

为简便计考虑同分布的情况,设 $X_1,\cdots,X_n$ 是从分布 F 中抽出的独立同分布样本.以 θ 记 F 的期望. T 统计量为

$$T_n = \sqrt{n}(\overline{X} - \theta) \Big/ \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2},$$

取随机加权均值统计量

$$H_n = \sum_{i=1}^{n} (X_i - \overline{X}) \xi_i.$$

这里 $\xi_1, \cdots, \xi_n$ 独立同分布，公共分布为 $\Gamma(4)$（且 $\xi_1, \cdots, \xi_n$）与 $(X_1, \cdots, X_n)$ 独立）. 易见

$$E^*(H_n) = 0, \quad \mathrm{Var}^*(H_n) = 4 \sum_{i=1}^{n} (X_i - \overline{X})^2.$$

于是

$$H_n / (\mathrm{Var}^*(H_n))^{1/2} = \frac{1}{2} H_n \Big/ \sum_{i=1}^{n} (X_i - \overline{X})^2 \equiv \widetilde{H}_n,$$

$\widetilde{H}_n$ 的条件分布（在给定 $X_1, \cdots, X_n$ 之下）已可以逼近 T_n 的分布. 为了得到更高的逼近阶，对 $\widetilde{H}_n$ 略加变换：定义

$$m_k = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^k, \quad k = 2, 3,$$

令

$$\omega(u) = u - \frac{1}{2\sqrt{n}} \frac{m_3}{m_2^{3/2}} u^2 + \frac{1}{16n} \frac{m_3^2}{m_2^3} u^3,$$

证明了：若存在 $\delta > 0$ 使 $\int_{-\infty}^{\infty} |x|^{6+\delta} \mathrm{d}F(x) < \infty$ 而 F 的方差不为 0，则有

$$\sup_y |P(T_n \leqslant y) - P^*(\omega(\widetilde{H}_n) \leqslant y)| = o\left(\frac{1}{\sqrt{n}}\right).$$

3. 线性回归模型

考虑线性回归模型

$$Y_i = x_i^\tau \beta + e_i, \quad i = 1, \cdots, n,$$

$\{x_i\}$ 为已知 p 维向量序列，$e_1, e_2, \cdots$ 独立同分布，有期望 0 和方差 σ^2.

设 $C^\tau \beta$ 为一可估函数，其 LS 估计唯一且等于

$$C^{\tau}\hat{\beta} = C^{\tau}S_n^{-}X_nY_{(n)},$$

其中 $X_n = (X_1 \cdots X_n)$，$Y_{(n)} = (Y_1, \cdots, Y_n)^{\tau}$，$S_n = X_nX_n^{\tau}$ 而 S_n^{-} 为 S_n^{-} 的任一广义逆. 有

$$\operatorname{Var}(C^{\tau}\hat{\beta}) = \sigma^2 C^{\tau}S_n^{-}C \equiv B_n^2,$$

记

$$T_n = (C^{\tau}\hat{\beta} - C^{\tau}\beta)/B_n,$$

需要模拟 T_n 的分布，结果可用于作 $C^{\tau}\beta$ 的区间估计. 为此目的，把 T_n 表为

$$T_n = \sum_{i=1}^{n} h_{ni}e_i/\sigma,$$

其中 $h_{ni} = C^{\tau}S_n^{-}X_i/(C^{\tau}S_n^{-}C)^{1/2}$. 以 $\hat{e} = (\hat{e}_1, \cdots, \hat{e}_n)^{\tau}$ 记残差向量，即 $\hat{e} = Y_{(n)} - X_n^{\tau}\hat{\beta}$. 用 $\sum_{i=1}^{n} \eta_i h_{ni}\hat{e}_i$ 模拟 $\sum_{i=1}^{n} h_{ni}e_i$，此处 $(\eta_1, \cdots, \eta_n) \sim D(4, \cdots, 4)$. 考虑到 T_n 的方差为 1，我们用

$$D_n = \sum_{i=1}^{n} \eta_i h_{ni}\hat{e}_i \Big/ \Big(\operatorname{Var}^{*}\Big(\sum_{i=1}^{n} \eta_i h_{ni}\hat{e}_i \Big) \Big)^{1/2}$$

去模拟 T_n 的分布. 这里如前，Var^{*} 是指在给定 $X_1, \cdots, X_n$ 的条件下的条件方差. 由于 $(\eta_1, \cdots, \eta_n)$ 与 $(X_1, \cdots, X_n)$ 独立，以及

$$\operatorname{Var}(\eta_i) = 5/[n(4n+1)],$$

$$\operatorname{Cov}(\eta_i, \eta_j) = -(3n+1)/[n^2(4n+1)], i \neq j.$$

再考虑到容易证明的关系式 $\sum_{i=1}^{n} h_{ni}\hat{e}_i = 0$，有

$$\operatorname{Var}^{*}\Big(\sum_{i=1}^{n} \eta_i h_{ni}\hat{e}_i \Big) = \frac{8n+1}{n^2(4n+1)} \sum_{i=1}^{n} h_{ni}^2 \hat{e}_i^{2}.$$

在一些条件之下证明了：以概率 1 成立

$$\sup_{y} |P^{*}(D_n \leqslant y) - P(T_n \leqslant y)| = o\Big(\frac{1}{\sqrt{n}} \Big).$$

这些条件包括：$\{x_i : i \geqslant 1\}$ 为有界列；S_n/n 当 $n \to \infty$ 时收敛于一正定方阵；$E|e_1|^3 < \infty$ 以及 e_1 的特征函数 $f(t)$ 满足 Cramer 条件，即

$\sup_{|t|\geq \varepsilon} |f(t)| < 1$ 对任何 $\varepsilon > 0$.

还讨论了有关统计量的分布渐近展开，得到了相应的结果.

五　最小一乘估计和 M 估计

考虑线性模型

$$Y_i = X_i^\tau \beta_0 + e_i, i = 1, \cdots, n, \tag{5.1}$$

其中各符号的意义与第四节一样，只是此处把回归系数向量记为 β_0.

β_0 的最小一乘估计，又称最小 L_1 模估计，简称 L_1 估计，定义为下面的极小值问题的解 $\hat{\beta}_n$：

$$\min_{\beta} \sum_{i=1}^{n} |Y_i - X_i^\tau \beta| = \sum_{i=1}^{n} |Y_i - X_i^\tau \hat{\beta}_n|. \tag{5.2}$$

对最简单的情况 $X_1 = X_2 = \cdots = 1$，位置参数 β_0 的 L_1 估计就是诸观察值 $Y_1, \cdots, Y_n$ 的样本中位数. 在这一特殊场合很清楚：为使 L_1 估计 $\hat{\beta}_n$ 为 β_0 的相合估计，e_i 的中位数 $\mathrm{med}(e_i)$ 必须为 0. 这一点对一般情况也成立.

从历史上说，L_1 方法的起源，与应用上最广的最小二乘法（L_2 方法）一样的早. 但在很长一段时间中，L_1 方法没有多大的进展. 原因一则在于其计算不易. 在现代计算机出现之前，求解(5.2)这样的极值问题是一个很繁重的工作. 再则，由于(5.2)的解 $\hat{\beta}_n$ 没有显式表达式，这使得不仅谈不上其小样本理论，大样本理论也很困难. 这种状况直到本世纪 50 年代才开始有所改变. 60 年代后期以来，L_1 方法的大样本理论逐渐受到一些学者的注意和重视，迄今已取得了不少的进展.

L_1 方法之所以引起学者和应用者的重视，在于近代研究证明：在不少情况下 L_2 方法的表现不理想，或者说缺乏稳健性. 而在这些场合下 L_1 方法的表现往往较佳. 例如，[40] 中指出，对计量经

济学(Econometrics)中的一些问题，往往不宜假定模型(5.1)中的随机误差 e_i 有有限的方差. 这个条件对 L_2 方法很重要，而对 L_1 方法则不重要.

下面介绍一下我们在 L_1 方法大样本理论方面所做的一些工作，及与此有关的工作.

首先是 L_1 估计的相合性问题. 关于这个问题，从 60 年代末到 70 年代前期已有些初步的工作，见[40]. 1978 年[41]中发表了关於 $\hat{\beta}_n$ 的渐近正态性的第一个证明. 如果成立，则自然得出有关 $\hat{\beta}_n$ 的相合性的结果. 可惜的是，这个证明被发现有本质性的错误(见[48]). 1982 年[42]中发表了一个证明，要求的条件很强，例如要求模型(5.1)中的 $e_1,e_2,\cdots$(假定为独立同分布的)的公共分布有连续密度，以及 $\sum_{i=1}^{n} X_iX_i^{\tau}/n \to$ 正定阵 V 当 $n\to\infty$. 另外，1982 年[43]中发表了一个非线性回归的 L_1 估计相合性的证明. 所要求的条件很强且过于复杂. 1983 年[44]中也发表了一个证明，但其中假定 $X_1,X_2,\cdots$ 是独立同分布的随机向量而不一定是常向量. 这个情况处理起来简单一些，且从某种意义上说，是 $X_1,X_2,\cdots$ 为已知常向量情况的一个推论.

我们关于 L_1 估计相合性的结果，主要总结在[45]中. 其中证明了：若在模型(5.1)中 $e_1,e_2,\cdots$ 独立(不必同分布)，$\mathrm{med}(e_i)=0$，且存在常数 $\triangle>0, C>0$，使

$$P(o<e_i<h)\geqslant ch\leqslant P(-h<e_i<o),\ o<h<\triangle. \tag{5.3}$$

记 $S_n=\sum_{i=1}^{n} X_iX_i'$，及

$$d_n=\max_{1\leqslant i\leqslant n} X_i^{\tau}S_n^{-1}X_i, \tag{5.4}$$

则有

1. 若 $d_n=o(1)$，则 $\hat{\beta}_n\to\beta_0$, in pr.

2. 若 $d_n = o(1/\log n)$，则 $\hat{\beta}_n \to \beta_0$，a. s.

3. 若 $d_n = o(1/n)$，则 $\hat{\beta}_n$ 在下述意义下以指数速度收敛於 β_0：对任给 $\varepsilon > 0$，有只依赖于 ε 而不依赖于 n 的常数 $c > 0$ 存在，使

$$P(\|\hat{\beta}_n - \beta_n\| \geqslant \varepsilon) = o(e^{-cn}).$$

由条件(5.3)当然推出各 e_i 的中位数都唯一地为0.这个条件比中位数唯一更进了一步，它要求有“足够多的”概率集中在0附近.如果 e_i 都有在某区间 $(-c,c)$ 上连续的密度函数 $f_i(t)$，则当 $\inf\{f_i(t): |t| < c, i = 1,2,\cdots\} > 0$ 时，条件(5.3)成立.有反例表明：即使在 $e_1, e_2, \cdots$ 独立同分布的情况下，条件(5.3)也不能减弱为中位数唯一.

与 L_2 估计的相合性相比，提出了两个与 L_1 估计相合性有关的问题.如所周知：若在模型(5.1)中，$e_1, e_2, \cdots$ 独立同分布，$Ee_i = o$ 而 $o < \mathrm{Var}(e_i) < \infty$，则 β_o 的 L_2 估计强、弱相合的充要条件都是

$$\lim_{n\to\infty} S_n^{-1} = o. \tag{5.5}$$

问题是：如果可以合理地认为在 L_1 估计中，上述条件 $Ee_i = 0$ 和 $0 < \mathrm{Var}(e_i) < \infty$ 大致相当于 $\mathrm{med}(e_i) = 0$ 及(5.3)的话，那么：

1. 在这些条件(即 $e_1, e_2, \cdots$ 独立同分布，$\mathrm{med}(e_1) = 0$ 及(5.3))下，L_1 估计 $\hat{\beta}_n$ 的强、弱相合是否等价(因在 L_2 估计中为等价：都等价于(5.5))？

2. L_1 估计 $\hat{\beta}_n$ 弱相合的充分条件可否减弱为(5.5)，且这个条件还是必要的？

我们对第1个问题作了否定的回答：通过一个反例(见[45])表明：在对 $\{e_i\}$ 的上述假定之下，有可能 $\hat{\beta}n$ 是 β_0 的弱相合估计，但并非强相合的.

对问题2的前一部分作了肯定的回答.实际上，[46]中对更广的M估计证实了这一点.

困难的是问题2的后一部分——条件(5.5)对 L_1 估计 $\hat{\beta}_n$ 的弱相合是否必要.有一切理由相信这一点，但到现在为止还未能给

出证明.但得出了若干有关的结果(见[47]).例如,证明了若 x_i 的维数为 1,或者若 $\{X_i\}$ 有界且 e_i 的公共分布在 0 的邻域内有密度 f,而 f 满足一定的条件,则条件(5.5)对于 L_1 估计的弱相合性是必要的.

至于 $\hat{\beta}_n$ 的强相合,问题就更困难.最近的研究表明:条件 $d_n = o(1/\log n)$ 中的数量级 $o(1/\log n)$ 已没有多少改进的余地:对任何常数 $c_n \uparrow \infty$,条件 $d_n = o(c_n/\log n)$ 已不再是充分的.当然,这不意味条件 $d_n = o(1/\log n)$ 为必要.因为可以存在不只依赖于 d_n 的更弱的条件.事实上,现在尚未找到任何比 $S_n^{-1} \to o$ 更强的必要条件,更不用说充分必要条件了.

另一个更重要的问题是 L_1 估计的渐近正态性问题.在一个最简单的特例——$X_1 = X_2 = \cdots 1$ 之下,问题成为样本中位数 $\hat{m}_n$ 的渐近正态性.早就知道,这在 $e_1, e_2, \cdots$ 独立同分布,其公共分布有中位数 0 并在 0 的一个邻域内有连续而大于 0 的密度 f 时成立,且

$$2f(0)\sqrt{n}(m_n - \beta_o) \to N(0,1)$$

受到这个结果的启发,人们猜想,当模型(5.1)中的 $e_1, e_2, \cdots$ 满足刚才指出的条件,而 $\{X_i\}$ 满足适当的条件时,应有

$$2f(0)S_n^{1/2}(\hat{\beta}_n - \beta_0) \to N(0, I_p). \tag{5.6}$$

最先企图证明这个结果的工作是[41].嗣后陆续发表了一些另外的证明(见[42],[43].可惜的是,这些证明都包含有本质的错误(反例见[48]).直到 1990 年,由我国几位统计工作者合作发表的论文[48],才在对 $\{X_i\}$ 最一般的条件下作出了(5.6)的严格证明.对 $\{X_i\}$ 的条件就是

$$\lim_{n\to\infty} d_n = 0, \tag{5.7}$$

其中 d_n 由(5.4)定义,由于这个条件在 L_2 估计的渐近正态性中是必须的(相应的 Lindeberg-Feller 条件),故有理由认为,对 $\hat{\beta}_n$ 的渐近正态性来说,条件(5.7)已无法再减弱,前引的一些工作中为证明(5.6),多引入条件 $S_n/n \to V > 0$.这个条件比(5.7)要强得多.

现在我们转到 M 估计. M 估计最初是 60 年代由 Huber 提出的. 1973 年他发表了一篇重要论文[49], 将 M 估计用于线性回归 (5.1). 此法在于选定一个适当的定义于 R^1 的函数 ρ, 而用下述极小值问题

$$\sum_{i=1}^{n} \rho(Y_i - X_i^\tau \beta) = \min \tag{5.8}$$

的解 $\hat{\beta}_n$ 去估计 β_0, 选 $\rho(u) = |u|$ 即得 L_1 估计, 而 $\rho(u) = u^2$ 则相应于熟知的 L_2 估计.

Huber 在[49] 中考虑了 ρ 在 R^1 为严格凸且足够光滑的情况, 在一些条件之下, 证明了 $\hat{\beta}_n$ 的相合性和渐近正态性. 在他所考虑的情况下, 问题(5.8) 的解与下述方程的解等价:

$$\sum_{i=1}^{n} \varphi(Y_i - X_i^\tau \beta) X_i = 0, \quad \varphi = \rho'. \tag{5.9}$$

以后, 有的学者径直从方程(5.9) 出发(选择适当的定义于 R^1 的函数 φ) 去定义 M 估计. 除非 φ 处处连续, 就无法保证有一个相应的函数 ρ. 因此, 这实在是另一种不同的 M 估计. 这方面有代表性的工作可参看[50].

这后一种定义 M 估计的方法缺乏前一种方法所具有的直观性. 更重要的, 方程(5.9) 的解的存在性, 即使在 φ 不太复杂时, 也没有保证. 我们所做的一项工作[51] 证明了这一点.

1987 年, Dodge 等在一项工作[52] 中取 $\rho(u) = (1-\delta)u^2 + \delta|u|$ (δ 为已知常数, $0 < \delta < 1$). 这个形式是意图将 L_1 方法和 L_2 方法"混合"起来. 除了 $u = 0$ 一点外, 都有

$$\rho'(u) = \varphi(u) = \delta \cdot \mathrm{sgn}(u) + 2(1-\delta)u.$$

他们就取这个 φ 去考虑(5.9) 的解, 并认为与(5.8) 等价, 而事实上, 即使在最简单情况 $X_1 = X_2 = \cdots = 1$ 之下, 相应的方程(5.9) 即

$$\delta \sum_{i=1}^{n}\mathrm{sgn}(Y_i-\beta)+2(1-\delta)\sum_{i=1}^{n}(Y_i-\beta)=0 \tag{5.10}$$

也不以概率 1 有解. 事实上,[51] 中证明了:若 $e_1,e_2,\cdots$ 独立同分布,公共密度为 f. 以 c 记使函数 $Q(a)=\delta E|Y_1-a|+2(1-\delta)E(Y_1-a)^2$ 达到最小之点(c 唯一). 若 f 在 c 点处连续且 $f(c)>0$,则

$\lim\limits_{n\to\infty}P$(方程(5.10)有解) $=(1-\delta)/[(1-\delta)+\delta f(c)]<1.$

因此,我们的观点是:这种形式的 M 估计并不可取,在计算上它也不见得比(5.8)更方便.

自 Huber 的工作[49]后,直接从(5.8)出发讨论 M 估计的文章,几乎只局限於 L_1 估计这个情况. 我国两位统计工作者在 1988 年发表的文章[53]讨论了 M 估计的相合性问题,对 X_i 为随机向量或为常向量的情况都得出了若干结果. 这个工作的一个特点是没有假定 ρ 是凸的. 就我们所知,这是迄今仅有的一篇未作这假定的文章. 具体结果涉及过多的细节,不在此引述了. 至于在(5.8)定义之下M 估计的渐近正态性,在 ρ 为凸的情况,近年已由白志东等的工作[54]完满地解决了. 在这项工作(其细节不在此引述)中,在假定 ρ 为凸(不必严凸)及其他很少的条件(包括(5.7))之下,证明了与(5.6)类似的渐近正态性定理. M 估计相合性充分条件问题已由[46]解决了. 另外值得一提的是,[54]中还解决了在 M 估计的基础上作线性假设检验的大样本理论问题.

M 估计研究的一个很重要的方面是其线性表示问题. M 估计研究的困难,在于它没有一个显式的解析表达式. 故人们希望用一个独立随机变量的线性组合去逼近它,使其剩余能尽量地小. 具有这种性质的线性组合,就称为 M 估计的线性表示. 如果剩余以概率 1(依概率)收敛于 0,则称为强(弱)表示. 线性表示问题的研究,就在于在尽可能少的条件下建立这种表达式,且使剩余项以尽

可能高的速度收敛于 0. 一旦建立了这种表达式，就可以利用独立和的丰富理论，去处理关于 M 估计渐近理论的深层次的问题.

这方面的研究始于 Bahadur[55]1966 年关于样本中位数强线性表示的工作. 1989 年，Babu[56] 考虑了一般线性模型的问题，但仍限于最小一乘估计的情况. 1992 年，Rao 与 Zhao[57] 将 Babu 的结果推广到一般的 M 估计，但限制函数 ρ 为凸的. 最近，陈希孺[58] 考虑了一般线性模型且 ρ 不必为凸的情况. 所得的线性表示，其剩余项在某些情况下达到了最好的收敛阶. 他利用这种线性表示，解决了 M 估计的重对数律，以及 M 估计的分布收敛於正态分布的一致性速度问题.

以上讨论限于线性模型的情况. 近年来，我国也有一些学者把 M 估计的方法用于某些特殊的非线性模型，例如半参数模型的估计问题. 如施沛德和李国英在[59] 中研究了形如

$$Y_i = X_i'\beta + g(T_i) + e_i, 1 \leqslant i \leqslant n,$$

其中非线性部分 g 是一个定义于$[0,1]$ 的，有一定的光滑性的函数. 他们用 M 估计去估计线性参数 β 及函数 g(用分段多项式逼近)，取得了较好的结果. 例如，在较弱的条件下证明了关于 β 的估计的渐近正态性.

毫无疑问，与 20 年前 Huber 发表其奠基性的工作[49] 时相比，线性模型 M 估计的理论已有了十分深刻和本质的进展，但有待解决的问题还不少. 我们认为，最重要的问题，也就是今后努力的主要方向，应当是在尽可能少的条件下，建立一个在 ρ 不必为凸时的 M 估计的完善理论. 目下有关 M 估计的问题，如相合性、渐近正态性等基本问题，解决得比较完善的，多限于 ρ 为凸函数的情况. ρ 非凸的情况虽有了不少结果，但条件大都不理想. 例如，对 ρ 的光滑性要求过高. 而在应用上一些重要的例子中，ρ 的一阶导数也不见得处处存在. 非凸情况下的 M 估计有其特别的重要性. 因为，引进 M 估计的一个重要理由就是为了改善估计的稳健性，而

这要求当 $|x| \to \infty$ 时，$\rho(x) \to \infty$ 的速度不能太快. 例如 ρ 有界，或 $\lim_{|x|\to\infty} \rho(x)/|x| = 0$. 这时 ρ 就不是凸的.

参 考 文 献

1. Huber, P. J., Projection pursuit, Ann. Statist. 13(1985) 435—475.
2. 成平，李国英，投影寻踪——一类新兴的统计方法，应用概率统计 2 (1986), 267--276.
3. 成平，李国英等，投影寻踪讲义，中国科学院系统科学研究所 (1986).
4. Pollard, D., Convergence of Stochastic Processes, Springer-verlag, New York (1984).
5. 张航，几类 PP 型检验统计量的性质，应用数学学报 12(1989) 82—95.
6. Shi, P. D. and Li, G. Y., A Class of PP L-statistics for test, Acta Math. Appl. Sinica, English Series Vol. 4, No. 2, 1991, 158—172.
7. Shi, P. D., A Class of PP R-statistics, Sys. Sci. & Math. Scis, 4 (1991) 84—96.
8. Shi, P. D. and Li, G. Y., Bootstrapping PP L-statistics for test, Sys. Sci. & Math. Scis. 4(1991) (to appear).
9. Tang, X. L. and Li, G. Y., PP U-statistics and their applications, in "Nankai Subseries of Lecture Notes on Probability and Statistics" World Scientific, 1992.
10. 张航，PP 拟合优度检验，系统科学与数学，8(1988), 234—242.
11. 蔡越虹，用投影寻踪、自助法进行多元分布函数的拟合优度检验，系统科学与数学，11(1991) 52—62.
12. Li, G. Y. and Zha, W. X., PP Neyman test for multivariate goodness of fit, in "The development of Statistics: recent contributions from China, Pitman Research Notes in Mathematics Series 258". X. Chen et al eds Longman Academic Scientific & Technic, 1992.
13. Fridman, T. H., Exploratory projection pursuit, J. Amer. Statist. Assoc, 82(1987), 249—266.

14. 张健,成平, PP 检验的渐近功效,系统科学与数学, 9(1989) 370—382.

15. Alexander, K. S., Probability inequaties for empirical processes and a law of the iterated lorgarithm, Ann. Prob. 12(1984) 1041—1067.

16. Aleder, R. J. and Brown, L. D., Tail behaviour for suprema of empircal processes, Ann. Prob. 14(1986) 1—30.

17. Alder, R. J. and samorodnitsky, Tail behaviour for the suprema of Gaussian processes with applications to empirical processes, Ann. Prob. 15(1987). 1332—1351.

18. Öhrvik, J. Structure in noise, Technique Report, Dept. of Statistics. University of Stockholm(1987).

19. Öhrvik, J. On the distribution of the Kolmogorov distance in the worst direcetion, Technique Report, Dept. of statistics. University of Stockholm (1988).

20. Huber, P. J. Spurious Structure? Technical Report, Center for Intelligent Control Systems MIT (1988).

21. Sun, J. Y. The P-values of projection pursuit, Technical Report, No. 104, Dept. Statist. Stanford University.

22. 张健, PP Kolmogorov-Smirnov 统计量极限分布下界, Chinese Bulletin of Science 35 (1990) No. 19.

23. 张健,成平, PP Kolmogorov-Smirnov, 统计量其分布,尾部大样本上界,数学学报 34(1991) 388-402.

24. 朱力行,张健,成平,一类经验过程一致偏差的指数界及其应用,应用数学学报 16 卷 4 期 1993 年,545—559.

25. 朱力行,半空间下标集 P 桥上确界的尾概率及其在经验过程中的应用,科学通报 36(1991) 656—657.

26. 朱力行, Kolmogorov 距离在最坏方向上分布的概率估计,科学通报 35 (1990) 1625—1627.

27. 朱力行,以半空间为指标的经验过程的尾部性质,应用数学学报 16 卷 3 期 1993,420—430.

28. Zhang, J., The tail estimation for the suprema of P-bridge, Systems

Science and Mathematics Vol. 5, 1992, 274－286.

29. Zhang, J. Zhu, L. X. and Cheng, P., Exponential bounds for the uniform deviation of an empirical processes and its applications (1), J. Multivariate Anal. Vol. 47, No. 2, 1993, 250－268.

30. Cheng, P. and Zhu, L. X., The tail probability inequalities for the PP type Cramér -von Mises statistics in "Nankai subseries of lecture Notes on probability and statistics (1992)", World Scientific, 46.

31. 华罗庚,王元,数论在近似分析中的应用,科学出版社 (1978).

32. Wang Yuan and Fang Kaitai, Number theoretic method in applied statistics, Chin. Ann. of Math. 11 B : 1 (1990) 51—65.

33. Wang Yuan and Fang Kaitai, Number theoretic method in applied statistics. (11), Chin. Ann. of Math. 11 B : 3 (1990) 384—394.

34. Efron, B., Bootstrap methods another look at the jackknife, Ann. Statist. 7(1979), 1—26.

35. 郑忠国,随机加权法在线性模型中的应用,系统科学与数学, 8(1988), 151—166.

36. Zheng Zhongguo, Random weighting method, Advance in Mathematics 18(1984) 44—54.

37. Zheng Zhongguo, Random weighting method for T statistics, Acta, Mathematica Sinica, New Series 5(1989) 87—94.

38. 郑忠国,涂冬生,线性模型中的随机加权逼近,中国科学, A 辑(1988)6 期 561—575.

39. Rubin, D. B., The Bayesian buotstrap, Ann. Statist. 9(1981) 130—134.

40. Taylor, L. D. Estimation by minimizing the sum of absolute errors, in "Frontiers in Econometrics". New York, Academic Press (1974).

41. Basseet, G. Jr. & Koenker, R. Asymptotic theory of least absolute error regression, J. Amer. Statist. Assoc. 1978, 618—622.

42. Amemiya, A. Two stage least absolute deviations estimates, Econometrika, 1982, 689—711.

43. Oberhofer, W. The consistency of nonlinear regression minimizing the

L_1-norm, Ann. Statist. 1982, 316-319

44. Bloomfield, P. & Steiger, W. L. Least Absolute Deviations. Birknaser, 1983.
45. Chen Xiru, Zhao, Lingcheng, Bai Zhidong and Wu Yuehua, On consistency of minimum L_1-norm estimates in linear regression models. to appear in "The development of statistics: recent contributions from China Pitman Research Notes in Mathematics Series 258". Chen el al eds, Longman Academic Scientific & Technic, 1992.
46. Zhao, L. C., Rao, C. R. & Chen, X., A note on the consistency of M-estimates in linear models, in "Stochastic Process", 359—367, Springer-Verlag, 1993.
47. Chen Xiru, Zhao Lincheng and Wu Yuehua, On Conditions of consistency of ML_1 N estimates, Statistica Sinica. Vol. 3, No. 1, 1993, 9—18.
48. 陈希孺,白志东,赵林城,吴月华,线性模型中最小一乘估计的渐近正态性,中国科学,A辑 1990 年 5 月 449—463.
49. Huber, P. J. Robust regression, Ann. Statist. 1973, 799—821.
50. Yohai, V. J. & Maronna, R. A., Asymptotic behavior of M-estimators for the linear model, Ann. Statist. 258—268.
51. Bai, Z. D. Wu, Y. Chen, X. R. Miao, B. Q., On Solvability of an equation arising in the theory of M-estimates, Communications of Statistics, Theory and Methods 1990, 363—380.
52. Dodge, Y. & Jureckova, J., Adaptive combination of least squares and least absolute deviations estimates, in Statistical Data Analysis Based on L_1-norm and Related Methods. North Holland (1987).
53. Chen, X. R. and Wu, Y., Strong consistency of M-estimates in linear models, J. Multivariate Anal. 1988, 116—130.
54. Bai, Z. D., Rao, C. R. & Wu, Y., Unified theory of M-estimation of linear regression parameters, Tech. Report. No. 89-04, Penn. State University, 1989.
55. Bahadur, R. R., A note on quantiles in large samples, Ann. Math.

Statist. . 1966, 577—580.

56. Babu, G. J. Strong representation for LAD estimates in linear models, Probab. Theory Related Fields, 83, 547—558.

57. Rao, C. R. and Zhao, L. C. , Linear representation of M-estimates in linear models, Canadian J. statist. , 1992, 359—368.

58. 陈希孺,线性回归M估计的线性表示,中国科学23卷12期,1993年12月,1264—1275.

59. 施沛德,李国英,部分线性模型参数的M估计的渐近正态性,将在科学通报发表.

八五数学重点项目简介

Brief Introduction To 8 · 5 Major Projects of Mathematics

许忠勤
（国家自然科学基金委员会数理科学部）
Xu Zhongqing
（Department of Mathematical and
Physical Sciences, National Natural
Science Foundation of China）

八五数学重点项目简介
Brief Introduction To 8·5 Major Projects of Mathematics

"现代数学若干基本问题的研究"作为国家自然科学基金委员会"七五"重大项目已于1991年结题，1991年5月在南开数学研究所举行的第二次21世纪中国数学展望学术会议上进行了总结和验收. 本项目选择了现代数学中10个课题进行研究，这些课题一般都不限于单一学科，具有较强的综合性，在数学研究中都有重要意义. 它们既有很强的理论性，又有广泛的应用前景，大多数都是国际数学研究的热门领域，符合数学发展的趋势. 这个项目的组成在一定程度上也反映我国数学发展的现状和水平.

为了高质量地完成这项研究任务，我们对承担这项任务的成员做了精心的选择，把国内一些学术水平高、研究能力强的数学家，特别是一些优秀的中青年数学家吸收进来，承担这项研究任务. 例如，在承担项目研究的全部71名成员中，有学部委员12人，博士导师55人，正教授、正研究员67人. 这71人中，有33人先后获得国家自然科学奖26项(建国以来数学获得自然科学奖总数为39项)，获得科学院或部委级奖的还有18项. 正是由于对课题和参加的人员的选择做得较好，三年来本项目的研究工作取得了巨大的成绩. 在验收会上，专家小组对此做了充分的肯定，指出："本项目经过三年实施. 取得了巨大的成绩，胜利地完成了原定任务. 开创了若干中国独创的处于国际领先水平的研究领域；涌现出众多具有国际先进水平的优秀科研成果；形成了不少在国际上具有一定影响力的研究群体；培养了一批优秀的青年数学人才. 这些都将对中国今后的数学事业的发展产生深远的影响."

数学七五重大项目取得的成绩和这几年我国数学各方面取得的显著成就，以及广大数学家表现出来的团结进取、努力拼搏的精神，得到了中央和各方面领导的高度赞扬和充分肯定. 财政部两次给数学拨专款，国家自然科学基金委员会设立了数学天元基金. 天元基金的设立弥补了数学基金的不足，大大缓解了数学研究上经费的困难. 这样，一方面可以使面上的自由申请项目的资助强度有

较大的提高，另一方面使我们可以更好地安排八五期间的重大重点项目，全面地考虑数学各分支的发展.

我们从 1989 年下半年就开始酝酿、规划和安排八五期间数学的重大重点项目.

一方面，我们积极组织力量，努力争取数学中那些科学意义重大、应用前景广泛、研究力量雄厚的领域进入国家科委组织的重大项目. 经过与有关数学家的共同努力，现在"大规模科学与工程计算的方法和理论"，"非线性科学"这两个包含有数学的相当多方面的研究内容的重大项目已在国家科委通过实施，有 50 名左右的数学家获得高强度的资助，承担这两个重大项目的研究任务. 还有一个以数学为重要研究内容的项目"机器证明及其应用"，将作为国家科委的第二批重大项目通过立项.

另一方面，我们在数学界广泛地征求意见，反复地进行讨论，准备在国家自然科学基金委员会内组织一批数学的重点项目. 现在，有 24 项已经通过立项. 它们是：

(1) 解析数论；

(2) 代数数论；

(3) 代数几何；

(4) Lie 群及其表示理论；

(5) 整体微分几何及物理应用；

(6) 代数拓扑和微分拓扑；

(7) 经典复分析；

(8) 多复变的全纯映照理论；

(9) Euclid 群上的调和分析；

(10) 算子理论和算子代数；

(11) 非线性分析；

(12) 动力系统与 Hamilton 系统；

(13) 常微分方程的分支问题与多项式系统理论；

(14) 偏微分方程的一般理论；

(15) 椭圆型偏微分方程；

(16) 非线性发展方程；

(17) 粒子系统与随机场；

(18) 随机分析及其应用；

(19) 多维数据的统计理论；

(20) 应用统计；

(21) 计算复杂性的理论和应用；

(22) 随机与分析参数系统控制的数学理论；

(23) 优化的算法与理论；

(24) 组合数学.

这 24 个项目中，现在已有 18 个项目已通过实施. 下面对每个项目逐一地作些简单的介绍.

(1) 解析数论

解析数论是数论的一个主要分支，以 20 世纪 20 年代到今天，一直是研究十分活跃、成果极为丰富的领域. 在这一分支中有许多著名问题，如 Riemann 猜想、Goldbach 猜想、Lindelöf 猜想、素数分布、Waring 问题等. 对这些问题的研究和所取得的成果，已经对数论和其他数学分支的发展起了重大的推动作用. 但是，这些问题至今仍未最终解决.

解析数论的研究方法主要有复变函数论方法、指数和与特征和估计法、圆法、筛法. 其中最重要的是指数和与特征和的估计法，这个方法可以被应用在许多数论问题上，如 Waring 问题、Goldbach 猜想问题、素数分布问题等. 它是本项目要采用的主要方法.

我国在解析数论方面的研究有很强的实力，这是我国数学各分支中极少数能在国际上保持领先地位的一个领域. 继续开展这个方面的研究，保持我国在这个领域中的领先地位是十分重要的，

承担本项研究的是我国数论界最好的专家，如陈景润、王元、潘承洞都是国际上有名望的数论专家. 本项目研究的具体内容是：

① 改进三素数定理的常数；

② 改进一些重要问题：如 Waring 型问题、Goldbach 型问题、素数分布问题、无平方因子数分布问题、小区间中整数的最大素因子下界的估计问题等的结果；

③ 改进算术数列中的最小素数；

④ 改进特征和的估计；

⑤ 研究同余方程的最小解.

(2) **代数数论**

代数数论是现代数学中重要又活跃的分支，它与数学的其他分支联系非常广泛，并在数值计算、通讯网络和保密通讯等许多领域中有非常深刻和非常广泛的应用. 我国在这些方面的研究开始于 1980 年左右，但是进步非常快，现已取得了一批引起国际同行赞扬的达到国际先进水平的成果.

本项目研究的一些问题是七五重大项目“现代数学中若干基本问题的研究”中代数数论课题研究的继续. 这些问题都是当今代数数论中的前沿课题. 承担本项目的是我国在这个分支中的学术带头人或主要骨干，是一批基础好、能力强、水平高的中年数学家. 本项目具体的研究内容是：

① 研究二次域的精细结构. 包括确定特殊小类数实二次域，建立更多的二次域的类数公式，研究 Pell 方程猜想；

② 研究分圆数域与分圆函数域及其子域的类群和单位群的 Galois 结构；

③ 研究模形式理论及其应用. 包括对 Siegel-Jacobi 形式空间维数的计算和 Hecke 算子迹公式的计算，利用模形式研究二次型表数问题、格的性质及树码的构造等；

④ 研究某些椭圆曲线的 BSD 猜想及山群.

(3) **代数几何**

代数几何是基础数学的一个极重要的分支，它主要研究代数曲线、曲面及更高维的代数流形，与现代数学的几乎所有重要分支都有密切的关系，是当今世界上最热门的研究领域.

代数几何的一个基本问题是研究结构和分类问题(如曲面的结构、奇点的结构、Abel 簇的分类等). 本项目研究的一些课题是七五重大项目“现代数学中若干基本问题的研究”中代数几何课题研究的继续. 承担本项目的大多数是青年数学家，他们的工作在各方面大都处于国际上的前沿地位或领先地位，他们是我国在代数几何领域中研究的学术带头人或主要骨干. 特别是肖刚，他是世界公认的纤维化方法的专家，他的文章一般都在学术水平最高的杂志上发表. 本项目的主要研究内容是：

① 运用代数曲面的纤维化方法解决代数曲面理论中的一些重要问题. 如一般曲面的自同构群和基本群；小亏格的纤维化的结构及分类；纤维扭曲理论的应用等；

② 群概形作用论和 Abel 簇的结构；

③ 某些次数较小的正规射影曲面的性质以及奇点的分布，这类曲面的分类；

④ 射影簇的刻划，射影簇的陈示性数的不等式研究以及 Frenet 标架理论；

⑤ 高维参量空间理论研究.

(4) **代数拓扑和微分拓扑**

拓扑学是基础数学的重要分支. 低维流形是代数拓扑中最重要的研究领域，在流形的拓扑学中，二、三、四维流形的研究与五维以上流形的研究在问题和方法上都有很大差别. 20 世纪 70 年代中期以来，由于创造了多种新方法，国际上低维流形的研究取得了重大突破，成为基础数学中最活跃的核心领域之一. 流形的浸入和嵌入问题一直是微分拓扑中最重要的研究领域. 另外，几何应用也很

广泛，从50年代后就很快发展起来，也是国际上非常活跃的研究领域.

本项目主要研究低维流形和流形的浸入与嵌入. 其中低维流形是七五重大项目“现代数学中若干基本问题的研究”中低维流形研究的继续. 承担本项目的都是我国一些卓有成就的中年数学家，尤其是姜伯驹教授和李邦河教授的研究工作在国际上有较大的影响，是国内拓扑学的学术带头人. 本项目具体的研究内容是：

① 三维流形及其映射的研究，包括三维流形间映射度非零的映射、纽结上的剜补术、Reidemeister 挠不变量与纽结理论、不动点理论的关系等；

② 微分流形的在给定同伦类中的浸入与嵌入的研究. 包括存在性与分类、法协边群的计算等；

③ 经典的和广义的 Adams 谱序列的研究，包括谱 $V_{(4)}$ 和 $V_{(n)}$ 的存在性，用 Adams 谱序列确定一些同伦群的非零元素；

④Steenrad 代数上不稳定模的几何实现研究.

(5) 整体微分几何及其应用

整体微分几何是现代数学的重要分支，它和分析、拓扑、代数等分支相互渗透、相互促进，与理论物理也有极密切的关系. 如物理学中的 Yang-Mills 理论，在数学中就是纤维丛上的联络理论. 理论物理中的非线性 σ 模型，在数学上是一类重要的调和映照. 近几十年来，整体微分几何无论是在理论研究上还是在方法上都取得了一些重大成就，成为基础数学中最活跃、最核心的领域之一.

本项目要研究的一些课题，都是微分几何中的重要和热门问题，也是七五重大项目“现代数学中若干基本问题的研究”中整体微分几何及其物理应用课题研究的继续. 承担本项目的胡和生教授等是我国在微分几何研究的学术带头人和主要骨干，他们的研究工作在国内外都有一定的影响. 本项目主要的研究内容是：

①Minkowski 空间和 Lorentz 流形的曲面理论；

② 孤立子的几何理论，如孤立子方程的形式、高维孤立子的几何性质等；

③ 调和映照的理论和性质，如存在性、不存在性、正则性、稳定性等；

④ 子流形的理论，包括极小子流形、等参子流形、Dupin 子流形等；

⑤ 低维流形上的 Yang-Mills 理论及其性质，Witten-Jones 不变量的计算；

⑥ 具有几何背景的非线性分析和低维拓扑等问题的几何方法研究；

⑦ 仿射极大曲面、常仿射平均曲率曲面以及相关的流形上的分析问题的研究，复仿射微分几何的研究.

（6）经典复分析

复分析的理论是 Cauchy 在 19 世纪 20 年代开始建立起来的，经过 Riemann、Weirstrass 等数学家的发展，已成为数学的一门非常重要的分支. 它既有十分完美的理论，又有非常广泛的应用，它与数学的各个分支都有密切的联系，也为其他学科分支提供有效的数学理论和方法.

本项目主要研究经典复分析中的三个基本问题：值分布论、复动力系统、拟共形映照及其应用. 这几个问题虽都有多年的历史，但由于它们的重要性和新的发展，在国际上仍有不少数学家在研究. 我国在这几个方面都作出过一些有价值的贡献，有较强的实力. 这几方面的研究领域也是七五重大项目“现代数学中若干基本问题的研究”中复分析课题研究的继续. 承担本项目的杨乐教授等是我国在复分析领域中的学术带头人和主要骨干，他们的研究工作在国际上有较大的影响. 本项目的研究内容是：

① 对亚纯函数及其导数的总亏量作精确的估计，并考察及其相关的 Frank 猜想、Goldberg 猜想、Hayman 不等式的最佳系数，

研究函数值分布与迭代、不动点等的相互制约与联系；

② 研究亚纯动力系统的基本问题，Julia 集的基本性质、游荡稳定域的存在性及构造问题等；

③ 带参数的复解析动力系统：πe^2，$\lambda \tan Z$ 等的结构稳定性和遍历理论，Julia 集的测度问题，Mondebrot 集与分形几何的关系；

④ 高维 Möbius 群与拟共形映照的基本性质及其在复解析动力系统中的应用、高维 Möbius 群的不变量表示理论等；

⑤ 具有非零亏格的 Riemann 面上的 Riemann 边值问题及在弹性力学的接触问题和断裂问题上的应用.

(7) **调和分析**

调和分析是当前实分析发展的主流方向. 80 年代以来，Euclid 空间上的调和分析出现了令人瞩目的新进展. 表现在：① 第三代奇异积分算子的研究为非线性分析开拓了一条新道路；② 小波分析方法的出现，把光滑性与衰减性高度和谐地统一起来，人们可以利用小波基更好地刻划和研究各种函数空间；③ 流形上调和分析的发展，为调和分析中曲率和度量等几何特征有关问题的解决开阔了前景；④ 非紧 Lie 群的无穷维表示，将对称空间上的调和分析与代数数论、代数几何、算子代数和量子化中曾是互不相关的概念都联系起来，以 Toeplitz 算子和 Hankel 算子研究为代表，将把现代数学作为一个有机整体来研究. 总之，Euclid 群上的调和分析将出现新的突破. 本项目的研究确有重大科学意义.

承担本项目的是我国在调和分析领域中的学术带头人和主要骨干，有很强的研究能力和很高的学术水平，在国际上也有一定的影响. 本项目的研究内容是：

① 改进小波基的构造，使之适合各种理论与实际的需要；

② 利用小波基刻划各种函数空间；

③ 研究奇异积分算子. 包括振荡积分，带粗糙核的奇异积分，球形和算子以及非线性逼近算子；

④ 调和分析中与曲率有关的问题；

⑤ 非紧 Lie 群的无穷维表示与 Hankel 算子.

(8) **常微分方程的分支理论与多项式系统理论**

研究常微系统在扰动下的变化规律，是结构稳定还是结构不稳定即出现分支现象，这在理论和应用两个方面都是至关重要的. 对结构稳定性的问题，国外以 Smale 为代表，国内以廖山涛为代表进行了大量的研究，取得了丰硕的成果. 对结构不稳定即分支现象的研究，由于力学、物理、生物、化学、医学、工程技术以及经济学的需要也成为国际上的热门领域，在近 20 多年发展异常迅猛. 我国在这方面的研究起步较晚，从 70 年代末才开始，但也有一些相当好的工作，而且从事这方面研究的人越来越多.

本项目提出的一些问题，有的是分支理论的基本问题，有的是多项式系统中未被解决的重要问题，都有重要的意义，承担本项目研究的都是我国常微分方程界的学术带头人和主要骨干，有很好的基础. 本项目具体的研究内容是：

① 全局分支和局部分支. 包括余维大于 2 的奇异向量场分支，某些 Hopt 分支以及高维流的同宿环、异宿环的分支的研究；

② 各种分支的稳定性研究；

③ 向量场的规范形理论、计算和应用；

④ 某些二次系统极限环的唯一性、唯二性问题及轨线图的全局拓扑结构；

⑤ 将二次系统定性研究的一些重要结果推广到 n 次多项式系统；

⑥ 把分支理论和方法应用于多项式系统的研究.

(9) **偏微分方程的一般理论**

偏微分方程一般理论的研究在 60 年代末、70 年代初以来获得蓬勃的发展，一个重要的标志是微局部分析理论的形成和发展. 本项目提出的一些问题，都是偏微分方程理论中的重要问题，既有重

要的几何和物理背景，又有重要的理论意义．大多是国际偏微分方程领域研究的热门课题．

本项目确定的几个研究课题大多是七五重大项目“现代数学中若干基本问题研究”中非线性分析课题研究的继续．承担本项目研究的成员都是国内这些领域的学术带头人或主要骨干，有很强的研究能力．本项目具体的研究内容是：

① 非线性微局部分析的理论和应用．包括仿微分算子理论、高维非线性方程解的奇性结构、非线性方程边值问题的一般理论、高维激波中奇性干扰问题，非线性奇性传播问题等；

② 经典散射理论和 Fractal 物体上的散射问题；

③ 广义 Fuchs 型算子、全特征算子与非线性亚椭圆算子．

（10）**椭圆型偏微分方程**

椭圆型偏微分方程是偏微分方程的基本和重要内容之一，它的理论在近百年来的数学发展中占有相当重要的位置．无论是正则型的方程还是退化型的方程都有很多重要的几何和物理背景．因此，不仅在理论上而且在应用上都有重要的意义．对它们的研究虽已有很长的历史，仍一直是热门的研究领域．

我国在这个领域的研究有较好的基础和自己的特色，有一支较强的研究队伍．本项目要研究的一些问题大多是七五重大项目中非线性分析课题研究的继续．承担本项研究的成员都是我国知名的数学家，是本学科的学术带头人，他们有很高的学术水平．具体的研究内容是：

① 完全的非线性椭圆型偏微分方程的研究．包括深化与拓广 Caffarelli 关于粘性解的局部估计方法；蜕化 Monge-Ampere 方程的解的存在性与正则性；非散度型蜕化拟线性方程的粘性解；等等；

② 以渗流力学为背景研究散度型蜕化拟线性抛物型方程组，多维情况解的结构、唯一性、最佳正则性、奇异解和非奇异解的存

在性等等；

③ 研究除标准球面 S^n 以外，其他典型的流形（如球面和复射影空间）上极小曲面的构造等.

(11) **粒子系统和随机场**

随机场和粒子系统是现代概率论与统计物理相互渗透中产生的新的数学分支. 近些年它在国外发展很快，已成为概率论研究的中心内容之一. 它的深入研究有可能为物理、化学、生物、经济、计算机网络等提供新的工具和模型，同时极大地促进数学本身的发展. 因此，它的重大科学意义是很明显的.

本项目研究的主要内容是七五重大项目"现代数学中若干基本问题的研究"中粒子系统和无穷维分析课题研究的继续. 我们在这个研究领域的研究有一定的基础，如在 Markov 过程方面做出过很好的工作，在粒子系统方面，针对非平衡物理中一些著名模型提出了无穷维反应扩散过程的模型，取得一批有我国特色的成果. 承担本项目研究的是我国在这个领域的学术带头人和主要骨干. 具体的研究内容是：

① 找出无穷粒子 Markov 过程一般的存在唯一性定理；

② 研究一些无穷粒子系统的遍历性、非遍历性、动力学极限及大偏差等问题；

③Gibbs 随机场的性质、相变及其极限理论（包括某些非中心极限定理及大偏差等）；

④ 在神经网络的粒子系统模型中，研究实现记忆、学习、归纳、演绎、识别、指挥命令等功能；

⑤ 研究两指标 Markov 过程的分析性质和轨道性质等；

⑥ 用求解随机微分方程、超过程泛函、二参数转移函数等方法构造 Markov 场及表示；

⑦ 研究超过程泛函的分析性质、位势理论；

⑧ 开展随机过程及随机过程中的 Fractal 性质研究.

（12）**优化的算法和理论**

最优化是应用数学的重要分支，随着计算机的发展和普及，优化的算法和理论迅速发展，目前已成为国际上最活跃的研究领域。最优化的研究对现代经济管理和工程设计中的最优决策起着重要的作用，这种作用在于为最优决策分析中的定量化提供科学的依据和方法。因此，它有非常广阔的应用前景。

线性和非线性优化与组合优化是最优化中重要的研究领域，也是本项目研究的主要内容。这些内容也是七五重大项目“现代数学中若干基本问题研究”中最优化课题研究的继续。承担本项目研究的，大都是我国该领域的学术带头人，都有很好的研究工作。本项目具体的研究内容是：

① 改进和完善线性规划中内点算法和理论，创造新的内点算法，并推广到非线性优化中去。由此寻求某些组合优化的新算法，研制新算法的相应软件；

② 非线性规划的理论和算法的研究。包括对拟牛顿法、直接方法的进一步研究；非线性优化一般算法模型的收敛性研究；连续性算法的复杂度理论的研究；

③ 组合优化的理论和算法研究。包括多面体理论的研究；对几种重要的组合优化问题如旅行售货商问题、时间表问题进行算法分析，提出更好的近似算法；离散事件动态系统、柔性制造系统及图论中一些极值问题的基本性质研究。

以上12个项目已作为数学天元基金的重点项目于1991年1月开始实施。

（13）**多复变的全纯映照理论**

多复变函数论是近几十年迅速发展起来的综合性、交叉性很强的学科分支。它和分析、代数、几何、拓扑等学科彼此渗透，互相促进，成为现代数学中的主流学科之一。我国在这个领域的研究有较好的基础，华罗庚教授及其学生在有界对称域方面的工作处于

国际领先的地位，被称为中国学派. 70 年代末，在华先生的学生陆启铿教授领导下，研究的领域在不断扩大.

本项目着重研究全纯映照下各种几何量的变换、构造与估值；全纯映照下不变的线性、非线性偏微分方程的整体解；以及全纯映照理论在共形场论中的应用. 这些都是多复变函数论的前沿问题，对这些问题的深入研究，将对其他分支的研究产生重要的影响. 这些问题也是七五重大项目“现代数学中若干基本问题的研究”中复分析课题研究的继续. 承担本项目研究的陆启铿教授等是我国在多复变领域中的学术带头人和主要骨干，在国际上有较大的影响. 本项目具体的研究内容是：

① 研究双全纯映照下 C^n 中的域按曲率条件的分类问题；

② 微分形式的热方程的解及其性质与应用的研究；

③ 复流形特征值估计及谱性质和 Pinching 问题；

④ 紧复的负曲率流形的分类；

⑤ 调和方程的特征根、Schwarz 引理，及 $C_1 > 0$ 时的 Calabi 猜想；

⑥ 复流形的 L^2 上同调与一般上同调的关系以及它们的应用；

⑦ 复结构的大范围形变与复流形的模问题；

⑧ 全纯向量丛的分解与分类；

⑨ 典型流形上的热核、特征值及格林形式；

⑩ 拟共形映照、Riemann 面上的模空间；

⑪ Bedford-Pinchuk 定理的推广；

⑫ 拟凸域上全纯不变量的比较估值以及非对称齐性域的自同构群.

(14) 非线性分析

非线性分析是近二十年来迅速发展起来的新的数学分支. 这是一个由微分几何、常微分方程、偏微分方程、拓扑学、泛函分析与经典分析融汇交织形成的分支. 它正不断以其获得的新成果显示

出强大的生命力，正是由于它在理论上和应用上的极端重要性，才吸引着越来越多的数学家、尤其是许多能力强思想活跃的青年数学家加入到它的研究行列，这又极大地扩大了其研究领域，推动着它更加突飞猛进的发展.

本项目着重研究与调和映射、哈密尔顿系统有关的分析问题以及其他物理、几何与经济中的变分问题. 这类问题因其自身的重要性以及遇到的困难带有的典型性，成为当今世界十分注目的研究课题，对这些问题的深入研究将对其他分支的研究和发展起重要的推动作用. 这些研究内容也是七五重大项目“现代数学中若干基本问题的研究”中非线性分析课题研究的继续. 承担本项目研究的张恭庆教授等是我国在非线性分析领域中的学术带头人和主要骨干，在国际上有较大的影响. 本项目具体的研究内容是：

① 与调和映射、极小曲面及预定中曲率曲面有关的方程；

② 预定曲率确定 Riemann 流形的方程；

③ 对带奇性的 Hamilton 系统，围绕三体问题研究周期轨、同宿轨、异宿轨；研究多体问题中解的碰撞问题，如碰撞次数、非碰撞以及碰撞的渐近行为；

④ 研究 Hamilton 系统的 Mather 集以及 Moser 极小 biliation 等问题；

⑤ 发展辛几何中的椭圆方法，Floer 同调，进一步研究 Arnold 猜测；

⑥ 微分包含及其在数理经济、微分方程中的应用.

(15) 多维数据的统计理论

数理统计是一门应用性很强的学科. 实际问题经常出现对多维数据的统计处理问题，解决这类问题难度很大，因此，多维数据的统计处理问题是近一二十年国际统计界最重要的热门研究领域之一.

本项目提出的几个问题，如 L_1 模及有关方法的理论、投影追

踪的理论和方法、回归分析等都是多维统计理论中的前沿问题，受到国内外统计界普遍的关心和重视. 我国在这几个领域的研究有一定的基础. 这几个问题也是七五重大项目“现代数学中若干基本问题的研究”中多维数据的统计推断课题研究的继续. 承担这个项目的都是我国统计界的学术带头人或主要骨干. 本项目的具体内容是：

① 回归系数最小 L_1 模估计和 M 估计的相合性问题；

② M 方法的检验和回归系统 M 估计的 Bahadur 表示；

③ M 方法的推广及在线性以外模型中的 L_1 和 M 方法；

④ 投影追踪统计量尾部概率和部分线性模型的估计；

⑤ 更一般的投影追踪回归，多维总体参数的估计检验；

⑥ 在投影追踪中的 Bootstrap 方法；

⑦ 可靠性综合的方法与理论和置信限的研究；

⑧ 随机加权法的研究.

(16) 随机分析及其应用

随机分析是国际概率论最活跃和最富于成果的分支之一，80年代发展起来的 Malliavin 分析，狄氏型及白噪声分析吸引了许多国际知名的概率论学家从事这些领域的研究，取得了重要进展. 这些理论不仅对概率论及随机过程理论的发展有着基本的重要性和深远影响，而且对数学其他分支（如微分方程、微分几何、位势理论）及量子物理有重要的应用. 此外，有关半鞅及其应用在国际上也继续受到重视.

本项目研究的几个主要课题是七五重大项目“现代数学中若干基本问题的研究”中粒子系统和无穷维分析课题研究的继续，承担本项目的是我国在随机分析方面的学术带头人和主要骨干，他们有很强的能力，都作出了很好的工作，在国际上也有一定的影响. 本项目具体研究内容是：

① 白噪声分析的理论及在量子物理中的应用；

② 狄氏型的理论及在量子物理中的应用；

③ 无穷维随机分析的一般理论；

④ 半鞅与随机分析的理论及在随机过程的极限理论、排队系统和随机过程统计中的应用；

⑤ 大偏差理论的系统化及在数学物理中的应用；

⑥ 高斯空间上的随机变分学和量子白噪声分析.

(17) **随机和分布参数系统控制的数学理论**

人类生活的许多领域都涉及到控制. 从工业过程、机器人到航天飞机、空间站等高技术的发展，控制理论都起着关键的作用. 而控制理论是以现代数学中若干理论为基础的，反过来控制理论的研究对数学的发展又提供了新的源泉和要求. 随机和分布参数系统控制理论是控制论中极重要、极活跃的领域，它与高技术关系密切，理论性强、难度大、待研究的问题多. 本项目研究的随机系统的辨识与控制，分布参数系统的控制和随机系统的最优过程控制是控制论中的三个基本问题，也是国际控制论界关心的热门课题. 对这几个问题的深入研究，无论对控制理论的发展和对在高科技中的应用都有十分重要的意义.

本项目研究的问题也是七五重大项目"现代数学中若干基本问题的研究"中辨识与控制课题研究的继续. 承担本项目研究的是我国控制论方面(数学部分)的学术带头人和主要骨干，都有很好的工作基础，在国际上也有一定的影响. 本项目具体的研究内容是：

① 连续时间随机系统 ELS 估计的控制问题和极点配置问题的研究；

② 柔性结构系统的性能、参数辨识、镇定及控制问题的研究；

③ 一般时变随机系统的适应控制和大范围求极值问题的研究；

④ 非线性偏微分方程描述系统的最优控制的存在性、必要条

件和解的正则性；

⑤ 具有跳跃测度的随机系统的性态和新理论研究；

⑥ 倒、正向随机微分方程解的存在性、唯一性、有界性及连续依赖性；

⑦ 随机控制系统的代数、几何理论、奇摄动方法和最大值原理；

⑧ 分布参数系统的概率方法，鲁棒反馈控制和稳健性分析.

(18) **算子理论和算子代数**

算子理论和算子代数是现代数学的重要分支，它的主要推动力是来自数学物理和量子物理中数学模型的需要，它的理论的形成和发展体现了代数、几何、分析的综合发展和相互渗透. 它对数学的其他领域特别是非交换几何、非交换拓扑以及 V. Jones 的指标理论起了很重要的作用，对本项目的研究有重要意义.

本项目包括算子理论和算子代数两部分内容，前者我国有一定的基础，后者是一个新的分支领域，涉及到 C^*-代数、K 理论、非交换微分几何及指标理论等. 这些都是国际数学界普遍感兴趣的问题. 我国在这个领域起步较晚，但也有一些较好的工作. 承担本项研究的是在泛函分析算子理论方面的学术带头人和主要骨干，有的在算子代数研究中有较强的实力. 具体的研究内容是：

① Hanker 算子的谱理论. 包括局部有界性、紧性和 SP 类的刻画；

② Toeplitz 算子代数分析；

③ Hilbert 模的刚性理论和局部化变换公式；

④ 一般线性算子理论的研究；

⑤ 多变量算子理论的等价条件和对偶定理；

⑥ 线性和非线性算子半群的理论和应用；

⑦ Von Neumann 代数的超自反性，Nest 代数的局部导数、自同构及插值问题等；

⑧ C^*- 叉积(Crossed products) 及 K- 理论的研究;

⑨ V. Jones 指标理论及其与核映射的关系;

⑩完全正线性映射和完全有界映射所确定的 C^*- 代数的同构及在 K- 理论中的应用.

以上 6 个项目已作为第一批数学重点项目于 1992 年 1 月开始实施.

从上述已经实施的 18 个项目看,无论是数学天元基金的重点项目还是数学重点项目都有以下共同的特点:

1. 研究内容新、交叉性强. 这些项目研究的课题几乎都是本分支领域的前沿课题,并且与其他一些分支有密切关系,这也反映当今数学发展的趋势. 这些项目的任何突破都将对本学科的发展乃至整个数学的发展产生重大影响.

2. 承担研究任务的数学家学术水平高、研究能力强,且以中青年为主. 在承担全部 18 项的 102 名人员中有博士导师 68 人,占数学全部博士导师的 1/3,其中学部委员 9 人,占数学全部学部委员 26 名的 1/3.

从年龄看,12 个天元重点项目平均年龄是 52.3 岁,6 个数学重点项目的平均年龄是 48.6 岁. 我们相信由这些人承担重点项目的研究在今后五年内能够做出好的工作.

(19) 动力系统和 Hamilton 系统

动力系统与 Hamilton 系统是现代数学的重要分支. 作为探索随时间演变的自然现象的规律及其复杂性的有力工具,它在物理、力学、生物、化学、气象、天文、工程控制及社会经济等学科有着广泛的应用. 它与微分方程、拓扑、微分几何、复分析和数学物理等分支有着深刻的内在联系. 因此,本项目有重要的理论意义和实际意义.

我国在动力系统方面的研究有较好的基础和较高的水平. 著名数学家廖山涛教授在 60 年代就从事这个方面的研究,他独辟蹊

径，首先提出阻碍集和典范方程组的概念，在结构稳定性及Ω稳定性的研究方面，做出了一系列高水平的工作，在国际上有很好的影响.

本项目提出的几个问题，都是动力系统和 Hamilton 系统中的基本问题，其中有些方面是七五重大项目"现代数学中若干基本问题的研究"中动力系统课题研究的继续. 具体研究内容是：

① 微分动力系统的扰动和稳定性问题；

② 拓扑动力系统的周期运动、熵、混沌等问题；

③ 周期 Dffing 方程的非线性现象；

④ 非线性 Hamilton 系统的周期解.

(20)Lie 群及表示理论

Lie 群及其表示理论是近 20 多年迅速发展起来的数学分支. 它和数学的其他分支如微分几何、拓扑、调和分析、多复变函数都互相影响、互相渗透、互相促进，它在理论物理、量子化学中有越来越多的应用，这种趋势还在不断发展. 因此，本项目的研究意义是重大的.

同计算机数学、数论等分支相比，我国在 Lie 群及表示理论方面的研究还比较薄弱. 近些年，随着国际上对 Lie 群的研究越来越热，国内一些中青年数学家开始从事这方面的研究，取得了一些好的结果. 把本项目列为重点项目，不仅是 Lie 群及表示理论自身发展的需要，也是其他相关学科如调和分析、微分几何、理论物理等发展的需要.

本项目所确定的几个方面研究内容都是国际上热门的研究课题，它们是：

① 约化 Lie 群酉表示及相关问题；

② 齐性空间微分几何；

③ Kac-Moody 代数；

④ Lie 群及表示理论在相关学科如数学物理、多复变函数、微

分方程等中的应用.

(21) **计算复杂性的理论和应用**

计算复杂性属于应用数学，它是研究计算的经济性的一个数学分支. 由于它在经济建设和军事技术的进步已经和正在发挥越来越大的作用，对计算机科学理论和基础数学本身(如代数几何、数论和微分几何等)的研究也将发挥重要作用，近20多年发展很快，美国国家研究委员会把计算复杂性的发展作为反映现代数学主要发展趋势的标志之一. 在我国计算复杂性的研究显然起步很晚，但已越来越受到各方面的重视，一些优秀的青年数学家在这个领域的研究工作相当出色. 如堵丁柱在1989年春成功地解决了Steiner树的最短程问题，论文发表后，在美国数学界引起很大反响.

本项目提出的几个方面的研究内容都是国际上这个领域的热门课题，且有非常明确的应用目标. 具体内容是：

① P,NP,PSPACE等复杂性类的性质和相互关系；

② 单向函数的存在性，因子分解问题的复杂性及其在密码学中的应用；

③ Boole函数电路复杂性及在神经网络中的应用；

④ 组合优化算法的复杂性、高效率非完整求解算法及在超大规模集成电路中的应用；

⑤ 连续性问题数值分析的计算复杂性.

(22) **应用统计**

应用统计是用数学方法从事数据的收集、分析、解释和处理的一门学科. 由于这些工作在工业、农业、国防、医学等许多领域都非常需要，甚至必不可少，世界各国都非常重视这一学科的发展. 过去十多年统计学取得了惊人的进步. 设备和通讯技术的发展提出了大量的复杂的数据，而计算机的出现和发展使计算能力空前提高，从而能处理各种复杂的庞大的数据. 因而统计学变得更有用武

之地了.

建国以来,我国的应用统计工作者在工业中的质量控制、产品开发,农业中的产量提高措施,国防上的核工业和宇航技术,以及医学、气象、经济、石油与资源的开发等方面都做出了很大的贡献.但是,我国应用统计的水平,无论是理论、方法还是应用的覆盖面,与世界先进水平还有不小的差距.需要大力提倡和推动这个分支的研究工作.

本项目提出的研究内容都是应用前景十分明确的课题,完成这些课题所用的数学方法,如多元分析、可靠性理论、时间序列分析、抽样调查、计算统计等,我国都有较好的基础.具体研究内容是:

① 影响和诊断的理论与方法;

② 无失效数据与参数设计的统计分析;

③ 非线性回归与自回归分析;

④ 不等概率与多阶抽样方法;

⑤ 任意连续分布随机变量和计算机抽样方法;

⑥ 非平衡数据与稀疏观测的谱分析与预报;

⑦ 成分数据的统计分析.

(23) 组合数学

组合数学是应用代数方法、分析方法、概率论方法以及拓扑等方法研究离散对象关系结构的一门数学分支.在计算机科学、信息科学和经济管理中,经常要遇到对各种对象的关系结构进行定量和定性分析.可见,对这个分支的研究意义非常重大.近些年来,组合数学,特别是组合对象的结构、组合设计、图的理论、计数问题等方面的研究非常活跃,发展很快,对计算机科学和信息科学的发展起着越来越大的作用.

国内组合数学的研究虽然起步较晚,但进步较快,这些年做了一些好的工作.本项目提出的几个方面的研究内容都是在国际上

引人注目在国内有一定基础的课题. 具体内容是:

① 经典组合学问题,如某些类型组合对象的构造性问题与计数问题;

② 各种组合设计问题;

③ 计算组合学问题;

④ 图的理论及计算问题;

⑤ 组合恒等式、组合不等式及级数反演问题;

⑥ 组合论中出现的重要特殊函数.

以上(19) — (23) 的 5 个项目已作为第二批数学重点项目通过立项,于 1992 年夏天组织评审,1993 年 1 月开始实施.

此外,我们还组织 5 个重点项目,连同"非线性发展方程"分别在 1992 年和 1993 年内立项和组织实施.

这样,在整个八五期间,数学将有 3 个重大项目,30 个重点项目. 这些项目的研究基本上代表了我国数学研究的现状和水平.

为了给承担这些项目研究的数学家创造良好的工作条件,我们除了给这些项目以较高强度的资助外,还将为承担重点项目的每位数学家配备一台计算机和一台打印机(通过申请). 希望通过这一有重大意义的步骤,彻底更新我国数学研究的手段,以加快我国数学事业发展的进程.

在七五期间经费十分困难的情况下,我国数学家都做出了巨大的成绩. 我们相信,在八五研究经费有了很大增加、研究条件有了极大地改善的情况下,我国数学事业一定能取得更大的成绩,我国数学研究的整体水平一定能大大地提高.

注 本文在 1992 年初完成. 现在情况已有很大变化,数学八五重点项目已全部开始实施. 除上述项目外,还有

(24) **群与代数的表示论;**

(25) **非线性发展方程;**

(26) **计算复杂性的理论和应用；**

(27) **点集拓扑的一些新方向；**

(28) **泛函微分方程及分支问题；**

(29) **某些非线性数值分析；**

(30) **图论．**

今后一两年，我们一方面将分期分批对已实施的重点项目进行中期检查和验收，同时组织数学界对九五期间数学优先资助领域进行研讨，在此基础上提出九五期间我们将资助的重大重点项目．

1994 年 4 月 7 日